

BU-707-M  
May 1980

STUDENTIZED RESIDUAL RULES FOR THE MULTIVARIATE REGRESSION

SINGLE-OUTLIER PROBLEM

Steven J. Schwager

Biometrics Unit, 337 Warren Hall, Cornell University  
Ithaca, New York 14853

Abstract

The multivariate regression model single-outlier problem is viewed as a decision problem with  $n+1$  possible actions. These are  $D_0$ : decide that there are no outliers, and  $D_i$ : decide that the  $i$ th observation is an outlier, for each  $i$ . The problem is invariant with respect to a natural group of transformations, and has a family of maximally invariant matrix-valued statistics. For mean slippage alternatives, each observation is associated with a hypothesized distance from the null model  $Y(n \times p) = XB + U$  if it is an outlier. For any set of  $n$  such distances, the class of decision rules of the following form is invariant admissible: for arbitrary  $f_1, \dots, f_n > 0$ , choose action  $D_0$  if  $\max_j f_j s_j < K$ ; choose action  $D_i$  if  $f_i s_i = \max_j f_j s_j \geq K$ ; where  $R_i$  is the  $i$ th row of the least squares residuals matrix  $R$ , and  $s_i = R_i (R'R)^{-1} R_i'$ . Similar results are obtained for the variance slippage case, and for mean slippage with several outliers in a common direction.

To be presented at a meeting of the Society for Industrial and Applied Mathematics (SIAM) in Houston, Texas, November 6-8, 1980.

## 1. Introduction

The problem of dealing with outlying or aberrant observations has been studied from many viewpoints. In a regression setting, Anscombe and Tukey (1963, p. 146) considered outliers to be "observations that have such large residuals, in comparison with most of the others, as to suggest that they ought to be treated specially." Much has been written about the effects of outliers on even the simplest statistical analysis. A single outlier can have tremendous impact on a sample mean or on a regression line fitted by least squares. In a one-way layout, an outlier can greatly increase the estimated variance, obscuring real differences among group means. The presence of several outliers can produce even more distortion.

To sensibly propose and compare procedures, one must know what information is sought from the analysis. As Kruskal (1960) and Gnanadesikan (1977, p. 272) have noted, an observation may be an outlier for one purpose but not for another. Two possible aims mentioned by David (1970, p. 170) are (a) to indicate whether outliers are present in the data, and (b) to identify those observations which are aberrant.

When it is known that at most one outlier is present, we will be able to address (a) and (b) simultaneously.

We will consider the multivariate regression problem with multiple regressors, so the model is

$$(1.1) \quad Y(n \times p) = X(n \times k)B(k \times p) + U(n \times p) \quad .$$

The observation matrix  $Y$  and the regressor matrix  $X$  are known. The coefficient matrix  $B$  is unknown, as is the error matrix  $U$ , which consists of  $n$  independent rows, each multivariate normal  $N(0, \Sigma)$ . Conditions to be stated in Definition 2.1 must also be satisfied. This model

fully specifies the structure of the data as long as no outliers are present. Any observation whose distribution departs from this model will be termed an outlier. For instance, assume that the data constitute a random sample from a normal  $N(\mu, \sigma^2)$  population, unless there are outliers. If all observations are independent, but  $k$  of them have distributions other than  $N(\mu, \sigma^2)$ , then the latter are outliers. In the two models most widely used to represent the existence of outliers, the last  $k$  observations are also normally distributed. Under the mean slippage model, also called model A, all observations have variance  $\sigma^2$ , but  $k$  of the means differ from  $\mu$ , and possibly from each other. Under the variance slippage model, also called model B, all observations have mean  $\mu$ , but  $k$  of the variances differ from  $\sigma^2$ , and possibly from each other. These models were proposed by Grubbs (1950) and Dixon (1950). They have been used and generalized by many others.

It may be known that at most one outlier is present. This situation, referred to as the single-outlier problem, is commonly analyzed as a decision problem with  $n+1$  possible actions. One of these,  $D_0$ , represents the decision that there are no outliers. The decision that  $D_0$  is incorrect and the  $i$ th observation is the outlier is denoted by  $D_i$ , for each  $i=1, \dots, n$ . For various versions of this and the related problem of slippage of one among  $n$  sampled normal populations, theoretically optimal decision rules have been found. Many are listed by David (1970, pp. 178-184), most of these being for the univariate case. A typical univariate result states that, under certain natural conditions, the probability of correctly identifying an outlier when one is present is maximized by the rule based on the maximum absolute studentized residual  $|v|$ : choose  $D_0$  if  $|v| < C$ ; choose  $D_i$  if  $|v| \geq C$  and the  $i$ th studentized residual has absolute value  $|v|$ .

Although many types of alternatives to (1.1) can be proposed, we will focus on two. Mean slippage alternatives occur when some observations have means not equal to the corresponding rows of  $XB$ . Variance slippage alternatives occur when some observations have variances  $\lambda_i^2 \Sigma$ , where  $\lambda_i^2 > 1$ . In each case, there may be one or more such observations, and several may display different mean shifts or variance inflation factors. Although more general families of alternatives would also be of interest, they present formidable mathematical barriers to the derivation of optimality results.

Expository treatments of the outlier problem, accompanied by many references, are found in David (1970), Doornbos (1966), and Grubbs (1969). Doornbos also discussed the history of outlier research, as did Anscombe (1960).

Karlin and Truax (1960) treated the multivariate common mean model ( $Y = \mu + U$ ) single-outlier problem with mean slippage as a multiple decision problem. Using a loss function that is essentially zero-one, they showed that any symmetric, affine invariant Bayes decision procedure is based on the magnitude of the largest squared studentized residual.

Ferguson (1961) examined the multivariate common mean model single-outlier problem with variance slippage. He showed that the symmetric Bayes rules found by Karlin and Truax under model A are also symmetric Bayes for model B. For the univariate regression model with normal errors,  $y = Xb + u$ , Ferguson considered the single-outlier problem with mean slippage. The null hypothesis  $H_0$  is that the model is correct, the alternative  $H_1$  that the mean of the  $i$ th observation's distribution differs from the model by  $a_i$ , for  $i = 1, \dots, n$ . Define the  $i$ th squared studentized residual as

$$V_i^2 = m_{ii}^{-1} R_i^2 / \sum_{j=1}^n R_j^2 ,$$

where  $R_i$  is the  $i$ th least-squares residual and  $m_{ii}$  is the  $i$ th diagonal element of the matrix  $M(n \times n) = I - X(X'X)^{-1}X'$ . A decision rule is given by: choose  $D_0$  if  $\max_j V_j^2 \leq K$ ; choose  $D_i$  if  $V_i^2 = \max_j V_j^2 > K$ . Ferguson proved that this rule is invariant admissible for the problem when  $|a_i|$  is proportional to  $m_{ii}^{-\frac{1}{2}}$ , and that it is Bayes with respect to a prior distribution giving equal weight to  $H_1, \dots, H_n$ .

Gnanadesikan (1977) discussed multivariate outliers from a data analytic viewpoint. An extensive survey of the outlier literature is found in Barnett and Lewis (1978).

The remainder of this paper is organized as follows. The slippage models of interest, outliers, and the mean slippage and variance slippage versions of the multivariate regression model single-outlier problem are defined in Section 2. Useful necessary and sufficient conditions for two standard assumptions about the regression structure are found in Section 3, and both versions of the single-outlier problem are shown to be invariant with respect to a group  $G$  of non-singular affine-like transformations of the data matrix. A family of maximal invariants  $T$  with respect to  $G$  is obtained in Section 4, and the distribution of  $T$  under both mean slippage and variance slippage alternatives of the single-outlier problem is derived. In Section 5, these distributions are used to obtain a class of decision rules invariant admissible for both versions of the problem, based on the magnitudes of the squared studentized residuals. This class is surprisingly large, containing many rules besides those based on the maximum studentized residual. In Section 6, the case of mean slippage alternatives with several outliers departing from the

null multivariate regression model in a common direction is examined, and decision procedures related to those of the single-outlier problem are found to be invariant admissible. These generalize Murphy's (1951) decision rules for the univariate, common mean model, aside from a minor difference in the invariance structure.

The decision rules of this paper are based on squared studentized residuals or the generalized version of them found in Section 6. Earlier results in this area, such as those of Ferguson (1961), Karlin and Truax (1960), and Murphy (1951), seem to include only those rules which are Bayes with respect to a prior distribution assigning equal probabilities to all alternatives except  $H_0$ , and specifying that slippage parameters  $\delta_i$  or  $\lambda_i^2$  are proportional to  $m_{ii}^{-1}$ . Outlier identification based on the magnitude of the largest squared studentized residual has been advocated on these grounds. Corollary 5.1 shows that this class of decision rules is invariant admissible in multivariate regression situations not previously investigated.

In both the mean slippage and variance slippage single-outlier problems, varying the slippage parameters  $\delta_i$  or  $\lambda_i^2$  produces the wider class of invariant admissible rules of Theorems 5.1 and 5.3. Since the parameters quantifying the slippage are not known in almost all practical cases, the rules of this wider class merit consideration. Theorems 5.2 and 5.4 show that this entire class of rules is invariant admissible no matter what the values of the slippage parameters, further justifying their serious consideration. Analogous comments can be made for the mean slippage problem with several outliers in a common direction, regarding the class of invariant admissible rules established in Theorem 6.1.

The question of whether to multiply univariate ( $p=1$ ) regression residuals by  $m_{ii}^{-\frac{1}{2}}$  before examining them to detect the presence of outliers has received recent attention, e.g. Cook (1977). The multivariate analogue of this issue is whether to use a rule based on  $V_i^2 = m_{ii}^{-1} R_i S^{-1} R_i'$  or one based on  $s_i = R_i S^{-1} R_i'$ . Corollaries 5.1 and 5.2 show that either type of rule is invariant admissible for both single-outlier problems.

## 2. Model Definitions

The multivariate regression model is fundamental to the discussion.

Definition 2.1: The multivariate regression model is specified by the matrix equation

$$(2.1) \quad Y = XB + U \quad ,$$

where  $Y(n \times p)$ ,  $X(n \times k)$ ,  $B(k \times p)$ , and  $U(n \times p)$  satisfy the following conditions:

$$(2.2) \quad p + k \leq n \quad ,$$

$$(2.3) \quad \text{rank}(X) = k \quad ,$$

$$(2.4) \quad \{u_i : i = 1, 2, \dots, n\} \text{ are independent, where } u_i \text{ denotes} \\ \text{the } i\text{th row of the matrix } U \quad ,$$

$$(2.5) \quad u_i \sim N(0, \Sigma) \quad \text{for all } i \quad .$$

The regressor matrix  $X$  is fixed and known, and the observation matrix  $Y$  is known after the data are collected. The regression coefficient matrix  $B$  and the  $p \times p$  error covariance matrix  $\Sigma$  are unknown, as is the error matrix  $U$ . The known scalars  $n$ ,  $p$ , and  $k$  represent the number of independent observations, the dimension or number of components of each observation, and the number of columns in the regressor matrix, respectively.

Condition (2.2) is introduced to insure the estimability of all unknown parameters, as discussed in Press (1972, p. 210).



The matrix  $X$  must be of full rank to make  $X'X$  non-singular. The unique solution of the normal equations will then be employed in constructing test statistics. This condition could be dispensed with; the discussion can be extended to cover  $X$  of less than full rank by introducing the parameter  $\text{rank}(X)$  and generalized inverses throughout. The details of reparametrization to a model of full rank are found in several standard sources, for example, Graybill (1961, p. 235). This matter, though not difficult, will not be discussed further here.

The independence of the error vectors  $\{u_i\}$  reflects the underlying assumption that the  $n$  rows of  $Y$  represent a set of independently selected observations. Conditions (2.4) and (2.5) together specify that the  $u_i$ 's are a random sample from a multivariate normal population with mean  $0(1 \times p)$  and covariance matrix  $\Sigma$ .

Definition 2.2: In the multivariate regression model of (2.1),  $Y_i$  and  $X_i$  will denote the  $i$ th rows of matrices  $Y$  and  $X$ , respectively. The following column vectors are of length  $n$ :  $e_i$  consists of a 1 in the  $i$ th positions and 0's elsewhere, for  $i=1, \dots, n$ ; and  $e_0$  has every entry equal to 0.

An alternative way of specifying the multivariate regression model, equivalent to the formulation of Definition 2.1, is to say that  $Y_1, \dots, Y_n$  are independent  $1 \times p$  random variables, each multivariate normal, with common covariance matrix  $\Sigma$  and means

$$(2.6) \quad E(Y_i) = X_i B \quad \text{for } i=1, 2, \dots, n.$$

This set of  $n$  equations can be written more compactly as

$$(2.7) \quad E(Y) = XB.$$

Conditions (2.2) and (2.3) must also be met.

The multivariate regression model of Definition 2.1 will be taken as the model of the null hypothesis; if there are no outliers, it is the model from which the data have been obtained. An observation will be called an outlier if it is a realization of a random variable which does not conform to the null model's specifications. To investigate the possible presence of outliers, alternative models of two forms are proposed. Model A, referred to as the mean slippage model, hypothesizes that the means of some observations differ from those specified by the null regression model. Model B, referred to as the variance slippage model, hypothesizes that the variances of some observations are larger than the variance specified by the null regression model.

Definition 2.3: The multivariate regression model with mean slippage is specified by

$$(2.8) \quad Y = XB + A + U \quad ,$$

where  $Y$ ,  $X$ ,  $B$ , and  $U$  are as in (2.1),  $A(n \times p)$  is an arbitrary matrix, and conditions (2.2)-(2.5) are satisfied.

An immediate consequence of equations (2.8) and (2.5) is

$$(2.9) \quad E(Y) = XB + A \quad .$$

Thus, the matrix  $A$  gives the displacements of the true means of the observations from the means stipulated by the null regression model. The distribution of the observation  $Y_i$  is multivariate normal  $N(X_i B + a_i, \Sigma)$ , where  $a_i$  denotes the  $i$ th row of  $A$ . Such a  $Y_i$  violates the null model if and only if  $a_i \neq 0$ . The null regression model is the special case occurring when  $A = 0$ . In dealing with outlier problems, the usual procedure is to assume that most of the rows of  $A$  are zero vectors. Additional information about the nature of  $A$  may be required for the analysis of this model. For example,  $A$  must not consist entirely of columns contained in the column space of  $X$ .

Definition 2.4: The multivariate regression model with variance slippage is specified by equation (2.1), conditions (2.2)-(2.4), and the condition

$$(2.10) \quad u_i \sim N(0, \lambda_i^2 \sigma^2) \quad \text{for all } i,$$

where  $\lambda_1^2, \lambda_2^2, \dots, \lambda_n^2$  is a set of arbitrary positive scalars.

Under this variance slippage model, the distribution of the observation  $Y_i$  is multivariate normal with mean  $X_i B$  and covariance matrix  $\lambda_i^2 \Sigma$ . Such an observation does not conform to (2.5) of the null model if and only if  $\lambda_i^2 \neq 1$ . The null regression model is a special case of the variance slippage model, occurring when  $\lambda_i^2 = 1$  for all  $i$ . For outlier problems, it is usually assumed that most of the  $\lambda_i^2$  are equal to 1, and that the rest are greater than 1. This is because, if  $\lambda_i^2 < 1$ ,  $Y_i$  will be distributed more closely about  $X_i B$  than under the null condition  $\lambda_i^2 = 1$ . We will not deal further with this situation, which has been aptly referred to as the presence of "inliers".

This discussion motivates the following definitions, one for each slippage model, of an outlier.

Definition 2.5: In the multivariate regression model with mean slippage, the observation  $Y_i$  is an outlier if  $a_i \neq 0$ , where  $a_i$  denotes the  $i$ th row of the matrix  $A$  in (2.8). In the multivariate regression model with variance slippage, the observation  $Y_i$  is an outlier if  $\lambda_i^2 > 1$  in equation (2.10).

Several basic definitions pertaining to the multivariate regression model will now be given.

Definition 2.6: Based on the multivariate regression model, define

$$(2.11) \quad \hat{B}(k \times p) = (X'X)^{-1}X'Y \quad ,$$

$$(2.12) \quad R(n \times p) = \hat{U} = Y - X\hat{B} \quad ,$$

$$(2.13) \quad M(n \times n) = I - X(X'X)^{-1}X' \quad ,$$

$$(2.14) \quad S(p \times p) = R'R \quad .$$

It is well known that  $\hat{B}$  is the least squares and the maximum likelihood estimator of  $B$ , and that  $R$  is the matrix of residuals. The symmetric, idempotent, positive semi-definite matrix  $M$  transforms  $Y$  into  $R$ .

Two assumptions will be made regarding the structure of the regression. These will be motivated by considering properties of univariate residuals. The first is that no residual will be allowed to have variance zero, for this would force it to be zero and so make it impossible to detect the corresponding observation's departure from the model. Second, no pair of residuals will be allowed to have correlation +1 or -1. An outlier in such a situation would produce equal effects in more than one residual, presenting obvious difficulty in determining which observation is the source of the trouble. An illuminating example is given by Anscombe (1960), who points out that in a  $3 \times 3$  Latin square, the nine residuals consist of three sets of three equal values. These conditions will be formally stated in terms of  $M$ , using the fact that  $\text{Cov}(R) = \sigma^2 M$  for univariate regression.

Assumptions 2.1: In all regression models to be considered, it is specified that these conditions hold:

$$(2.15) \quad m_{ii} \neq 0 \quad \text{for all } i \quad ;$$

$$(2.16) \quad m_{ii}m_{jj} \neq m_{ij}^2 \quad \text{for all distinct } i \text{ and } j \quad .$$

Important necessary and sufficient conditions for (2.15) and (2.16) will be given in Section 3.

A special case of the outlier problem occurs when at most one outlier is present. This reduction of the problem restricts the parameters of the multivariate regression slippage models, Definitions 2.3 and 2.4. For the mean slippage problem, at most one row  $a_i$  of the matrix  $A$  can be non-zero. For the variance slippage problem, at most one constant  $\lambda_i^2$  can be greater than 1, the rest being equal to 1.

These situations can be viewed as multiple decision problems. The goal is to decide whether one of the observations is an outlier and, if so, to identify that observation. The decision theoretic framework for examining the problem of one outlier will now be supplied.

A set of  $n+1$  hypotheses, denoted  $H_0, H_1, H_2, \dots, H_n$ , will be considered. Under the null hypothesis  $H_0$ , the observations  $Y$  are distributed according to the multivariate regression model, so there are no outliers. Under hypothesis  $H_i$ , for each  $i=1, \dots, n$ , point  $Y_i$  is an outlier, but all of the  $n-1$  remaining observations are distributed as the multivariate regression model dictates.

The action space is  $\mathcal{Q} = \{D_0, D_1, \dots, D_n\}$ . Here  $D_0$  represents the action "decide that there are no outliers," and  $D_i$  represents the action "decide that the  $i$ th observation is an outlier," for  $i=1, \dots, n$ .

The elements  $\theta$  of the state space have several components: the parameters  $B$  and  $\Sigma$  of the multivariate regression model; an entry  $i$  specifying which point is an outlier; and an entry describing the location (for mean slippage) or inflation factor (for variance slippage) of the outlier. In the mean slippage case, this last component of  $\theta$  will be

the  $1 \times p$  vector  $a_i$ . When there are no outliers, the index  $i$  assumes the value 0, and  $a_0$  is taken to be a zero vector. Naturally,  $a_i$  must be non-zero when  $i \neq 0$ . In the variance slippage case, the last component of  $\theta$  will be the scalar  $\lambda_1^2$ . With no outliers, again  $i$  is 0, and  $\lambda_0^2$  is taken to be 1. The choice of  $a_0 = 0$  or  $\lambda_0^2 = 1$  insures the uniqueness of  $\theta$ .

A zero-one loss function will be adopted. If there are no outliers, deciding that any observation is an outlier gives a loss of one. When there is an outlier, a loss of one results from either an incorrect decision on its identity or the decision that no outliers are present. Subject to certain conditions, this loss function declares a decision rule to be good if it maximizes the value of the collection  $\{\Pr(D_i|H_i), \text{ all } i\}$  in some sense. For example, in Bayes calculations, with prior distribution  $(p_0, p_1, \dots, p_n)$  on the component of  $\theta$  specifying the outlier, minimizing the Bayes risk is equivalent to maximizing  $\sum_{i=0}^n p_i \Pr(D_i|H_i)$ . This will be taken up in detail in Section 4.

**Definition 2.7:** The multivariate regression model, mean slippage single-outlier problem is specified by these elements:

(i) Hypotheses  $H_i: Y = XB + e_i a_i + U$ , where  $a_i(1 \times p) \neq 0$  if and only if  $i \neq 0$ , and conditions (2.2)-(2.5) hold, for  $i=0, 1, \dots, n$ .

(ii) Action space  $\mathcal{A} = \{D_0, D_1, \dots, D_n\}$ , where  $D_i$  denotes the decision to act as if hypothesis  $H_i$  is true.

(iii) State space  $\Theta_{SM} = \{\theta = (i, a_i, B, \Sigma) : i \in \{0, 1, \dots, n\} ; a_i (1 \times p) \neq 0 \text{ if and only if } i \neq 0; \Sigma > 0\}$  .

(iv) Loss function  $L(\theta, D_j) = L[(i, a_i, B, \Sigma), D_j] = 1 - \delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$  .

Definition 2.8: The multivariate regression model, variance slippage single-outlier problem is specified by these elements:

(i) Hypotheses  $H_0$ : Equation (2.1) and conditions (2.2)-(2.4) hold, and

$$u_j \sim N(0, \Sigma) \text{ for all } j .$$

$H_1$ : Equation (2.1) and conditions (2.2)-(2.4) hold,

$$u_j \sim N(0, \Sigma) \text{ for all } j \neq i, \text{ and}$$

$$u_i \sim N(0, \lambda_i^2 \Sigma) \text{ with } \lambda_i^2 > 1, \text{ for } i = 1, 2, \dots, n .$$

(ii) Action space  $\mathcal{A} = \{D_0, D_1, \dots, D_n\}$ , where  $D_i$  denotes the decision to act as if hypothesis  $H_i$  is true.

(iii) State space  $\Theta_{SV} = \{\theta = (i, \lambda_i^2, B, \Sigma) : i \in \{0, 1, \dots, n\} ; \lambda_0^2 = 1; \lambda_i^2 > 1 \text{ if } i \neq 0; \Sigma > 0\}$  .

(iv) Loss function  $L(\theta, D_j) = L[(i, \lambda_i^2, B, \Sigma), D_j] = 1 - \delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$  .

### 3. Invariance of the Single-Outlier Problem

Standard definitions and theorems on invariance will be invoked in dealing with these problems. A full treatment of these can be found in either Ferguson (1967) or Lehmann (1959), accompanied by much related material. Ferguson discusses the reasons for using invariant decision rules in very general situations; conditions favoring alternative approaches are also mentioned. One further definition is needed.

Definition 3.1 (Ferguson, 1961): A decision rule is invariant admissible if it is admissible in the set of invariant rules.

Any invariant decision rule must be a function of a maximal invariant under  $G$ . This situation can be exploited to our advantage. Assume that a problem is invariant with respect to a group  $G$ , and that a maximal invariant  $T$  with respect to  $G$  has been obtained. If it is desired to consider only the class of decision rules which are invariant under  $G$ , the equivalent class of decision rules which are functions of  $T$  may be considered instead.

From this point on,  $\mathcal{U}$  will denote the space of  $n \times p$  matrices. A group of transformations from  $\mathcal{U}$  into itself will now be introduced, and it will be shown that both the mean and the variance slippage single-outlier problems are invariant under this group. The transformation



$g_{C,K}$  operates on a data matrix  $Y$  by a non-singular (right) matrix multiplication, which replaces each row  $Y_i$  by  $Y_i C$ , followed by addition of an  $n \times p$  matrix  $XK$  to the result. The post-multiplication of  $Y$  by  $C$  will affect the scale and location of the underlying distribution of the data matrix, and the addition of  $XK$  will further alter its location.

Definition 3.2: Let  $\mathcal{Y}$  denote the set of all  $n \times p$  real matrices. Define the transformation  $g_{C,K}: \mathcal{Y} \rightarrow \mathcal{Y}$  by

$$(3.1) \quad g_{C,K}(Y) = YC + XK \quad ,$$

where  $C$  is a  $p \times p$  non-singular matrix, and  $K$  is an arbitrary  $k \times p$  matrix. The set of all such transformations will be denoted by

$$(3.2) \quad G = \{g_{C,K}: C \text{ is } p \times p; \det(C) \neq 0; K \text{ is } k \times p\} \quad .$$

Lemma 3.1: The set of transformations  $G$  in (3.2) is a group.

Proof: The proof is routine and will be omitted.

Before establishing the invariance of the single-outlier problem under  $G$ , we give useful necessary and sufficient conditions for the assumptions (2.15) and (2.16). The identifiability of  $\theta$  is then shown for the mean slippage and variance slippage problems.

Theorem 3.1: For  $M = I - X(X'X)^{-1}X'$ ,  $m_{jj} = 0$  if and only if  $e_j \in CS_X$ , where  $CS_X$  denotes the column space of  $X$ .

Proof: (i) Without loss of generality, take  $j = 1$ . Assume  $e_1 \in CS_X$ , so  $Xv = e_1$  for some vector  $v$ . Let  $S = (X'X)^{-1}$ . Then

$$\begin{aligned}
 e_1 &= Xv = XS[(X'X)v] \\
 &= XSX'(Xv) \\
 &= XS(X'_1, \dots, X'_n)e_1 \\
 &= \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} SX'_1 \quad .
 \end{aligned}$$

Equating the first elements of the left and right hand sides gives

$$X_1 SX'_1 = 1, \text{ or } m_{11} = 0 .$$

(ii) Again take  $j = 1$ , and assume  $m_{11} = 0$  . Then  $m_{1i} = 0$  for all  $i$ , since by the idempotence and symmetry of  $M$ ,

$$0 = m_{11} = \sum_i m_{1i} m_{i1} = \sum_i m_{1i}^2 \quad .$$

It follows from the definition of  $M$  that  $X_i SX'_1 = \delta_{1i}$ , the Kronecker delta. Direct evaluation shows that  $XSX'_1 = e_1 \in CS_X$  . QED

Condition (2.15), that  $m_{ii} \neq 0$  for all  $i$ , thus insures that no  $e_i$  can be in  $CS_X$  . The alternative hypotheses  $H_i$ ,  $i=1, \dots, n$ , of the mean slippage single-outlier problem contain slippage terms  $e_i a_i$  with  $a_i \neq 0$  . Consequently, should a single outlier be present, it cannot be incorporated into  $H_0$  merely by changing the entries of  $B$  .

Theorem 3.2: For  $M = I - X(X'X)^{-1}X'$  and  $i \neq j$ ,  $m_{ii}m_{jj} \neq m_{ij}^2$  if and only if  $ae_i + be_j \notin CS_X$  for all non-zero scalars  $a, b$  .

Proof: (i) Without loss of generality, take  $i=1$ ,  $j=2$ , and let  $S$  denote  $(X'X)^{-1}$  . Assume that  $ae_1 + be_2 \in CS_X$  for some  $a, b \neq 0$  . Then there exists a vector  $v$  with

$$\begin{bmatrix} a \\ b \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Xv = XSX'Xv = XS(X'_1, \dots, X'_n) \begin{bmatrix} a \\ b \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} aX_1SX'_1 + bX_1SX'_2 \\ aX_2SX'_1 + bX_2SX'_2 \\ \vdots \\ aX_nSX'_1 + bX_nSX'_2 \end{bmatrix} .$$

Equating the first and second elements of the first and last expressions gives

$$am_{11} + bm_{12} = 0 \quad ,$$

$$am_{12} + bm_{22} = 0 \quad .$$

Thus  $-a/b = m_{12}/m_{11} = m_{22}/m_{12}$ , implying that  $m_{11}m_{22} = m_{12}^2$ .

(ii) Again take  $i=1, j=2$ , and assume that  $m_{11}m_{22} = m_{12}^2$ . Set

$a = m_{12}/m_{11} = m_{22}/m_{12}$ , and define the vector  $f(n \times 1) = (a \ -1 \ 0 \ 0 \ \dots \ 0)'$ .

It suffices to show that  $Mf = 0$ , for this is equivalent to

$f = X(X'X)^{-1}X'f$ , which demonstrates the existence of a vector  $ae_1 - e_2$

$\in CS_X$ . We must show

$$am_{i1} - m_{i2} = 0 \quad \text{for all } i=1, \dots, n \quad .$$

This is trivial for  $i=1,2$ , since  $m_{12} = am_{11}$  and  $m_{22} = am_{12}$ . Using the notation of univariate linear models, let  $Y$  and  $R$  denote the vectors of observations and residuals, respectively. It is well known that  $R = MY$ , and that  $\text{Cov}(R) = \sigma^2 M$ . Hence the correlation between residuals  $r_1$  and  $r_2$  is

$$\rho(r_1, r_2) = \text{sign}(a) \quad ,$$

and  $r_2 = ar_1$  with probability 1. Using this relationship, we obtain

$$m_{i2}\sigma^2 = E(r_i r_2) = aE(r_i r_1) = am_{i1}\sigma^2 \quad \text{for all } i.$$

Thus  $m_{i2} = am_{i1}$ , as was to be shown. QED

Lemma 3.2: For the multivariate regression model, mean slippage single-outlier problem,  $P_\theta = P_{\theta'}$ , implies that  $\theta = \theta'$ .

Proof: Take  $\theta = (i, a_i, B, \Sigma)$  and  $\theta' = (j, a_j, B', \Sigma')$ , and let  $Y \sim P_\theta$ ,  $Y' \sim P_{\theta'}$ . If  $P_\theta = P_{\theta'}$ , it follows that  $E(Y) = E(Y')$ , that is,  $XB + e_i a_i = XB' + e_j a_j$ . Rearranging terms,  $X(B - B') = e_j a_j - e_i a_i$ . The right-hand side is in  $CS_X$ ; from conditions (2.15) and (2.16) and Theorems 3.1 and 3.2, it is not hard to show that  $i = j$ ,  $a_i = a_j$ , and hence  $B = B'$ . Equating the covariance matrices of the first rows of  $Y$  and  $Y'$  shows that  $\Sigma = \Sigma'$ , concluding the proof that  $\theta = \theta'$ . QED

Example 3.1: To see that condition (2.16) is necessary for identifiability, take  $p=1$  and

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}; \quad B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad i=2, \quad a_i=1; \quad B' = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \quad j=1, \quad a_j=1.$$

Substitution shows that  $XB + e_i a_i = XB' + e_j a_j = (4 \ 4 \ 6 \ 6)'$ . So, for any choice of  $\Sigma > 0$ ,  $\theta = (i, a_i, B, \Sigma)$  and  $\theta' = (j, a_j, B', \Sigma)$  are not equal, but  $P_\theta = P_{\theta'}$ .

Lemma 3.3: For the multivariate regression model, variance slippage single-outlier problem,  $P_\theta = P_{\theta'}$ , implies that  $\theta = \theta'$ .

Proof: Take  $\theta = (i, \lambda_i^2, B, \Sigma)$  and  $\theta' = (j, \lambda_j^2, B', \Sigma')$ , and let  $Y \sim P_\theta$ ,  $Y' \sim P_{\theta'}$ . It follows from  $P_\theta = P_{\theta'}$ , that  $XB = E(Y) = E(Y') = XB'$ , so

$B = B'$  . Equate covariance matrices row by row, and observe that either all rows have the same covariance matrix, or one row differs from the rest. Whichever is the case, it is simple to show that  $\Sigma = \Sigma'$ ,  $i = j$ , and  $\lambda_i^2 = \lambda_j^2$ , so  $\theta = \theta'$  . QED

Theorem 3.3: The multivariate regression model, mean slippage single-outlier problem is invariant under the group  $G$  .

Proof: Choose  $g_{C,K} \in G$ ,  $\theta = (i, a_i, B, \Sigma) \in \Theta_{SM}$  of Definition 2.7. Then  $Y \sim P_\theta$  means that  $Y = XB + e_i a_i + U$ , the rows of  $U$  are independent  $N(0, \Sigma)$  random variables, and conditions (2.2) and (2.3) hold. It follows that

$$\begin{aligned} g_{C,K}(Y) &= YC + XK \\ &= XBC + e_i a_i C + UC + XK \\ &= X(BC + K) + e_i (a_i C) + UC \end{aligned}$$

The rows of  $UC$  are independent  $N(0, C'\Sigma C)$  random variables, so

$$\begin{aligned} g_{C,K}(Y) &\sim P_{\overline{g_{C,K}}(\theta)}, \text{ where} \\ \overline{g_{C,K}}(\theta) &= (i, a_i C, BC + K, C'\Sigma C) \end{aligned}$$

Lemma 3.2 establishes the uniqueness of  $\overline{g_{C,K}}(\theta)$  . Thus the family  $\{P_\theta, \theta \in \Theta_{SM}\}$  is invariant under  $G$  . And, for any  $g_{C,K}$  in  $G$  and  $D_j$  in  $\mathcal{A}$ ,

$$\begin{aligned} L(\theta, D_j) &= L[(i, \dots), D_j] = 1 - \delta_{ij} \\ &= L[\overline{g_{C,K}}(\theta), D_j] \quad \text{for all } \theta \in \Theta_{SM} \end{aligned}$$

Letting  $\widetilde{g_{C,K}}(D_j) = D_j$ , this demonstrates the invariance of the loss function under  $G$  . In fact, it proves a slightly stronger result, which will now be stated. QED

Corollary 3.1: For all  $g_{C,K} \in G$ ,  $\theta \in \Theta_{SM}$ ,  $D_j \in \mathcal{D}$ ,

$$L(\theta, D_j) = L[\overline{g_{C,K}}(\theta), D_j] .$$

Proof: Contained in the proof of Theorem 3.3.

Theorem 3.4: The multivariate regression model, variance slippage single-outlier problem is invariant under the group  $G$ .

Proof: Choose  $g_{C,K} \in G$ ,  $\theta = (i, \lambda_i^2, B, \Sigma) \in \Theta_{SV}$  of Definition 2.8. Then  $Y \sim P_\theta$  means that  $Y = XB + U$ , the rows of  $U$  are independent random variables with  $u_j \sim N(0, \Sigma)$  for all  $j \neq i$ ,  $u_i \sim N(0, \lambda_i^2 \Sigma)$ , and conditions (2.2) and (2.3) hold. Of course, if  $i = 0$ , we have  $u_j \sim N(0, \Sigma)$  for all  $j$ . Then

$$\begin{aligned} g_{C,K}(Y) &= YC + XK \\ &= X(BC + K) + UC . \end{aligned}$$

The rows of  $UC$  are independent multivariate normally distributed random variables with mean 0, and for  $i \neq 0$ ,

$$(UC)_j = U_j C \sim N(0, C' \Sigma C) \quad \text{for all } j \neq i ,$$

and

$$(UC)_i = U_i C \sim N(0, \lambda_i^2 C' \Sigma C) .$$

When  $i = 0$ ,  $(UC)_j \sim N(0, C' \Sigma C)$  for all  $j$ .

Comparison with Definition 2.8 shows that

$$\begin{aligned} g_{C,K}(Y) &\sim P_{\overline{g_{C,K}}(\theta)}, \quad \text{where} \\ \overline{g_{C,K}}(\theta) &= (i, \lambda_i^2, BC + K, C' \Sigma C) . \end{aligned}$$

The uniqueness of  $\overline{g_{C,K}}(\theta)$  follows from Lemma 3.3, establishing the invariance of the family  $\{P_\theta, \theta \in \Theta_{SV}\}$  under  $G$ . The proof of the invariance of the loss function under  $G$  is exactly the same as in the mean

slippage case of Theorem 3.3 and Corollary 3.1. QED

Corollary 3.2: For all  $g_{C,K} \in G$ ,  $\theta \in \Theta_{SV}$ ,  $D_j \in \mathcal{A}$ ,

$$L(\theta, D_j) = L[\overline{g_{C,K}}(\theta), D_j] \quad .$$

Proof: Contained in the proof of Theorem 3.4.

In the proofs of Theorems 3.3 and 3.4, it was found that  $\widetilde{g_{C,K}}(D_j) = D_j$  for all  $g_{C,K} \in G$  and  $D_j \in \mathcal{A}$ , so a decision rule for the mean slippage (variance slippage) single-outlier problem is invariant under the group  $G$  of Definition 3.1 if

$$d[g_{C,K}(Y)] = d(Y) \quad \text{for all } Y \in \mathcal{Y}, g_{C,K} \in G \quad .$$

Since the multivariate regression model, mean slippage and variance slippage single-outlier problems are invariant under  $G$ , attention will be restricted to decision procedures which are invariant under this group. As noted earlier in this section, any such rule must be a function of a maximal invariant with respect to  $G$  .

4. A Family of Maximal Invariants

Given  $X$  and the data matrix  $Y$ , let  $D$  denote the  $n \times n$  diagonal matrix with  $n - k$  1's followed by  $k$  0's on the diagonal,

$$(4.1) \quad D(n \times n) = \begin{bmatrix} I_{n-k} & 0 \\ 0 & 0 \end{bmatrix} .$$

Then there exists an orthogonal  $n \times n$  matrix  $P$  such that

$$(4.2) \quad P'MP = D .$$

Let  $P^i$  denote the  $i$ th column of  $P$ , and define

$$(4.3) \quad \Phi_1[(n-k-p) \times n] = \begin{bmatrix} P^1, \\ \vdots \\ P^{n-k-p}, \end{bmatrix}, \quad \Phi_2(p \times n) = \begin{bmatrix} P^{n-k-p+1}, \\ \vdots \\ P^{n-k}, \end{bmatrix} .$$

Then  $P'M = DP' = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ 0 \end{bmatrix}$ , where the zero matrix is of dimension  $k \times n$ .

Define an  $n \times p$  matrix  $Z = P'MY$ . We can define the sub-matrices  $Z_1$  and  $Z_2$  of  $Z$  by

$$(4.4) \quad Z_1[(n-k-p) \times p] = \Phi_1 Y \quad \text{and} \quad Z_2(p \times p) = \Phi_2 Y .$$

Then

$$(4.5) \quad Z = P'MY = \begin{bmatrix} \Phi_1 Y \\ \Phi_2 Y \\ 0 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ 0 \end{bmatrix} .$$

The decomposition of  $Z$  into  $Z_1$  and  $Z_2$  will be used in Sections 5 and 6, where the distribution of  $Z$  is derived under various conditions.



Definition 4.1: Given  $X$  and  $Y$ , let  $M$  be defined by (2.13), and choose an arbitrary but fixed orthogonal matrix  $P$  satisfying (4.2). Let  $P^i$  denote the  $i$ th column of  $P$ . For  $\Phi_1$  and  $\Phi_2$  given by (4.3), restrict attention to the subset of the sample space  $\mathcal{Y}$  consisting of  $\{Y \in \mathcal{Y} : \Phi_2 Y \text{ is non-singular}\}$ . Define a matrix-valued statistic  $T(Y)$  of dimension  $(n-k-p) \times p$  by

$$(4.6) \quad T(Y) = \Phi_1 Y (\Phi_2 Y)^{-1} = Z_1 Z_2^{-1} \quad .$$

For any  $X$  there is considerable latitude in selecting  $P$ . As  $P$  varies over values satisfying (4.2), a family of maximally invariant statistics is generated. However, once a particular  $P$  is chosen in Definition 4.1, it is held fixed through all subsequent steps of analysis, both here and in later sections.

The subset of the sample space excluded from consideration in Definition 4.1 consists of all  $Y$  such that  $\Phi_2 Y$  is singular. This will be a set of measure zero in the situations to be investigated, and will therefore have no effect on the analysis.

Theorem 4.1:  $T(Y)$  is a maximal invariant with respect to the group  $G$ .

Proof: Note that  $MX = 0$ , so

$$0 = P'MX = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ 0 \end{bmatrix} X \quad .$$

Invariance will be shown first. Choose any  $g_{C,K}$  in  $G$ .

$$\begin{aligned} T[g_{C,K}(Y)] &= T(YC + XK) \\ &= \Phi_1 (YC + XK) [\Phi_2 (YC + XK)]^{-1} \\ &= \Phi_1 YC [\Phi_2 YC]^{-1} \\ &= \Phi_1 YC C^{-1} (\Phi_2 Y)^{-1} \\ &= T(Y) \quad . \end{aligned}$$

Maximality will now be shown. Let  $CX$ ,  $C1$ , and  $C2$  denote the column spaces of  $X$ ,  $\Phi_1'$ , and  $\Phi_2'$ , respectively. Then  $\dim(CX) = k$ ,  $\dim(C1) = n - k - p$ , and  $\dim(C2) = p$ . It is known that  $\Phi_1 X = 0$ ,  $\Phi_2 X = 0$ , and  $\Phi_1 \Phi_2' = 0$ , so Euclidean  $n$ -space  $R^n = CX \oplus C1 \oplus C2$ . Each vector  $y \in R^n$  can thus be written uniquely as

$$(4.7) \quad y = y_X + y_1 + y_2 \quad \text{with} \quad y_X \in CX, y_1 \in C1, y_2 \in C2 \\ = Xu + \Phi_1'v + \Phi_2'w, \quad ,$$

where  $u$ ,  $v$ , and  $w$  are column vectors of lengths  $k$ ,  $n - k - p$ , and  $p$ . Given that  $T(Y_1) = T(Y_2)$ , it must be shown that  $Y_2 = g_{C,K}(Y_1)$  for some  $g_{C,K}$  in  $G$ . Using (4.7) and proceeding column by column,  $Y_1$  can be expressed uniquely as

$$Y_1 = XU_1 + \Phi_1'V_1 + \Phi_2'W_1 \quad .$$

Similarly,  $Y_2$  can be expressed uniquely as

$$Y_2 = XU_2 + \Phi_1'V_2 + \Phi_2'W_2 \quad .$$

These expressions may be used to evaluate

$$\Phi_1 Y_1 = 0 + \Phi_1 \Phi_1' V_1 + 0 = V_1 \quad \text{for } i=1,2, \\ \Phi_2 Y_1 = 0 + 0 + \Phi_2 \Phi_2' W_1 = W_1 \quad \text{for } i=1,2 \quad .$$

Now  $T(Y_1) = T(Y_2)$  implies that  $\Phi_2 Y_1 = W_1$  and  $\Phi_2 Y_2 = W_2$  are non-singular and that  $V_1 W_1^{-1} = V_2 W_2^{-1}$ . Let  $C = W_1^{-1} W_2$ , so

$$W_2 = W_1 C \quad \text{and} \quad V_2 = V_1 C \quad .$$

Letting  $K = U_2 - U_1 C$ , and substituting these expressions into the formula for  $Y_2$ ,

$$Y_2 = X(K + U_1 C) + \Phi_1' V_1 C + \Phi_2' W_1 C \\ = Y_1 C + XK = g_{C,K}(Y_1) \quad .$$

QED

The distribution of  $T$  will be derived under the alternatives  $H_i$ ,  $i = 0, 1, \dots, n$  of the mean slippage single-outlier problem. The early results of this section hold under more general conditions. The multivariate regression model with mean slippage given in Definition 2.3 allows an arbitrary slippage matrix  $A$ , which need not be of the form  $e_i a_i$ . Everything up to Lemma 4.1 is applicable to this broader model.

The following definition will prove useful.

Definition 4.2: Let  $Z^*$  denote the  $np \times 1$  column vector obtained by concatenating the columns of  $Z$ , so

$$Z^* = (z_{11}, \dots, z_{n1}, z_{12}, \dots, z_{n2}, \dots, z_{1p}, \dots, z_{np}) \quad ,$$

and let  $Y^*$  and  $R^*$  denote the  $np \times 1$  column vectors obtained from  $Y$  and  $R$  by the same procedure. Let  $z_j (1 \times p)$  denote the  $j$ th row of  $Z$ ,  $P^j$  the  $j$ th column of  $P$ , as in (4.3), and  $P_i$  the  $i$ th row of  $P$ . The first  $n - k$  rows of  $Z$  comprise the sub-matrices  $Z_1$  and  $Z_2$  defined in (4.4). Define

$$(4.8) \quad W(p \times p) = Z_2 \quad \text{and} \quad Q(p \times p) = T'T + I \quad .$$

We begin by finding the distribution of  $Z = P'MY = P'R$ . The expectation of  $Z$  is

$$E(Z) = P'M E(Y) = P'M(XB + A) = P'MA = DP'A \quad ;$$

$Z^* = (I_p \otimes P')R^*$ , where  $I_p$  denotes the  $p \times p$  identity matrix, so

$$\text{Var}(Z^*) = (I_p \otimes P')(\Sigma \otimes M)(I_p \otimes P) = \Sigma \otimes D \quad .$$

Hence the last  $k$  rows of  $Z$  are degenerate at  $O(1 \times p)$ , while the first  $n - k$  rows are independently normally distributed, each with covariance matrix  $\Sigma$ . The joint probability density function of the first  $n - k$  rows of  $Z$  is thus

$$(4.9) \quad \begin{aligned} & g_{z_1, \dots, z_{n-k}}(z_1, \dots, z_{n-k}) \\ &= c \exp \left[ -\frac{1}{2} \sum_{j=1}^{n-k} (z_j - P^j'A) \Sigma^{-1} (z_j - P^j'A)' \right] \quad , \end{aligned}$$

where  $c = (2\pi)^{-(n-k)p/2} (\det \Sigma)^{-(n-k)/2}$  is a constant depending only on  $n-k$ ,  $p$ , and  $\Sigma$ .

It is immediate from (4.8) that  $Z_1 = TW$  and  $Z_2 = W$ , so

$$(4.10) \quad Z = \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} W \quad .$$

To make the change of variables from  $Z_1, Z_2$  to  $T, W$ , we must compute the Jacobian of the transformation. Formulas found in Press (1972, p. 45) give

$$\text{Jac} = |\det \partial(Z_1, Z_2) / \partial(T, W)| = |\det W|^{n-k-p} \quad .$$

Then the distribution of  $T$  is

$$(4.11) \quad f_T(T) = \int f_{T,W}(T, W) dW = \int g_{Z_1, Z_2}[Z_1(T, W), Z_2(T, W)] \text{Jac} dW \quad .$$

This distribution will be computed under alternatives  $H_i$ ,  $i=0, 1, \dots, n$  of the multivariate regression model, mean slippage single-outlier problem. First, we derive several needed results.

Lemma 4.1: For the multivariate regression model, mean slippage single-outlier problem, under  $H_i$ ,  $i=0, 1, \dots, n$ ,

$$\sum_{j=1}^{n-k} z_j \Sigma^{-1} z_j' = \text{tr}[W \Sigma^{-1} W' Q] \quad ,$$

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} z_j' = a_i \Sigma^{-1} W' (T' \ I \ 0) P_i' \quad ,$$

and

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} A' P^j = m_{ii} a_i \Sigma^{-1} a_i' \quad .$$

Proof: For the first result, noting that  $Z'Z = W'QW$  follows easily from (4.10), we have

$$\sum_{j=1}^{n-k} z_j \Sigma^{-1} z_j' = \text{tr} \left[ \sum_{j=1}^{n-k} z_j' z_j \Sigma^{-1} \right] = \text{tr}[Z'Z \Sigma^{-1}] = \text{tr}[W'QW \Sigma^{-1}] .$$

For the second result, recalling that  $A = e_i a_i$ ,

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} z_j' = \text{tr} \left[ \sum_{j=1}^{n-k} A \Sigma^{-1} z_j' P^j \right] = \text{tr}[A \Sigma^{-1} Z' P'] = a_i \Sigma^{-1} Z' P_i' .$$

Substitution from (4.10) completes the proof. For the last result, note that

$$\sum_{j=1}^{n-k} P^j P^{j'} = (PD)(PD)' = M ,$$

and that  $A \Sigma^{-1} A'$  consists entirely of zeroes except for the value in the ith position, which is  $a_i \Sigma^{-1} a_i'$ . Then

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} A' P^j = \text{tr} \left[ \sum_{j=1}^{n-k} P^j P^{j'} A \Sigma^{-1} A' \right] = \text{tr}[M A \Sigma^{-1} A'] = m_{ii} a_i \Sigma^{-1} a_i' .$$

QED

From (4.8), (4.10), (4.11), and Lemma 4.1, we obtain the distribution of  $T$  under  $H_i$  as

$$(4.12) \quad f_T(T) = (2\pi)^{-(n-k)p/2} (\det \Sigma)^{-(n-k)/2} \alpha_i \int \exp[-\frac{1}{2} \text{tr}(W \Sigma^{-1} W' Q)] \\ + a_i \Sigma^{-1} W' (T' I O) P_i'] |\det W|^{n-k-p} dW ,$$

where

$$\alpha_i = \exp[-\frac{1}{2} m_{ii} a_i \Sigma^{-1} a_i'] .$$

We proceed to derive a more useful expression for  $f_T(T)$  by a method due to Karlin and Truax (1960, Sec. 9). Define  $\tau(p \times 1) = \Sigma^{-\frac{1}{2}} a_i'$  and  $E(p \times p) = \Sigma^{-\frac{1}{2}} W' Q \Sigma^{-\frac{1}{2}}$ , which is permissible since  $Q$  is positive definite and symmetric whenever  $T$  exists. A change of variables from  $W$  to  $E$  shows that

$$(4.13) \quad f_T(T) = (2\pi)^{-(n-k)p/2} (\det Q)^{-(n-k)/2} \alpha_{i\rho} ,$$

where

$$\rho = \int \text{etr}[-\frac{1}{2}EE' + EQ^{-\frac{1}{2}}(T' \ I \ 0)P_i'\tau'] |\det E|^{n-k-p} dE .$$

Define the matrix  $N(p \times p) = \tau P_i \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} Q^{-1} (T' \ I \ 0) P_i' \tau'$ . Then, by Vinograd's Theorem (Press, 1972, p. 37), there exists an orthogonal  $p \times p$  matrix  $K$  for which

$$Q^{-\frac{1}{2}}(T' \ I \ 0)P_i'\tau' = KN^{\frac{1}{2}} .$$

Making the variable transformation  $F = EK$  gives

$$\rho = \int \text{etr}[-\frac{1}{2}FF' + FN^{\frac{1}{2}}] |\det F|^{n-k-p} dF .$$

Let  $L(p \times p)$  be the orthogonal matrix which diagonalizes the matrix

$$\Lambda(p \times p) = LN^{\frac{1}{2}}L' .$$

Then  $\Lambda$  is of rank one, since  $N^{\frac{1}{2}}$  is of rank one, and  $L$  can be chosen to make the first diagonal element of  $\Lambda$  positive and the rest equal to zero. Letting  $\lambda_{ii}$  denote the  $i$ th diagonal element of  $\Lambda$ , we can express this as  $\lambda_{11} > 0$  and  $\lambda_{ii} = 0$  for all  $i \neq 1$ .

Now make a final transformation from  $F$  to the  $p \times p$  matrix variable  $G$  defined by  $G = LFL'$ . Then  $\rho$  can be expressed as

$$\begin{aligned} \rho &= \int \text{etr}[-\frac{1}{2}GG' + G\Lambda] |\det G|^{n-k-p} dG \\ &= \int \exp[-\frac{1}{2}\text{tr}(GG') + g_{11}\lambda_{11}] |\det G|^{n-k-p} dG . \end{aligned}$$

We now obtain a more convenient expression for  $\lambda_{11}$ .

Definition 4.3: The scalars  $s_i$ ,  $\delta_i$ , and  $\alpha_i$  are defined by

$$(4.14) \quad s_i = P_i \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} Q^{-1} (T' \ I \ 0) P_i' ,$$

$$(4.15) \quad \delta_i = a_i \Sigma^{-1} a_i' = \tau' \tau, \quad \text{and}$$

$$(4.16) \quad \alpha_i = \exp[-\frac{1}{2} m_{ii} \delta_i] .$$

It is easy to show that  $\text{tr}[(vv')^{\frac{1}{2}}] = (v'v)^{\frac{1}{2}}$  for any column vector  $v$ . Proceeding to the evaluation of  $\lambda_{11}$ , we observe that

$$\lambda_{11} = \text{tr}(\Lambda) = \text{tr}(L' \Lambda L) = \text{tr}[(s_i \tau \tau')^{\frac{1}{2}}] = s_i^{\frac{1}{2}} \text{tr}[(\tau \tau')^{\frac{1}{2}}] = s_i^{\frac{1}{2}} \delta_i^{\frac{1}{2}} .$$

We have proved

Theorem 4.2: For the multivariate regression model, mean slippage single-outlier problem, the distribution of  $T$  under  $H_i$ ,  $i=0,1, \dots, n$ , is

$$(4.17) \quad f_T(T) = (2\pi)^{-(n-k)p/2} (\det Q)^{-(n-k)/2} \alpha_i \rho(s_i, \delta_i) ,$$

where  $\rho(s_i, \delta_i)$  is defined by

$$(4.18) \quad \rho = \rho(s_i, \delta_i) = \int \exp[-\frac{1}{2} \text{tr}(GG') + g_{11} s_i^{\frac{1}{2}} \delta_i^{\frac{1}{2}}] |\det G|^{n-k-p} dG .$$

The distribution of  $T$  will now be derived under the alternatives  $H_i$ ,  $i=0,1, \dots, n$ , of the variance slippage single-outlier problem. The invariance of  $T$  under the group  $G$  allows us to reduce the algebraic complexity of the calculations by taking  $B=0$  and  $\Sigma=I$  without loss of generality. The elements of  $Y$  are then independent normally distributed random variables, each with expected value zero. Under hypothesis  $H_0$ , each of them has variance one. Under  $H_i$ ,  $i=1, \dots, n$ ,

$$(4.19) \quad \text{Cov}(Y_j) = I_p \quad \text{for } j \neq i, \quad \text{Cov}(Y_i) = \lambda_i^2 I_p,$$

where  $Y_j$  is the  $j$ th row of  $Y$ .

Definition 4.4: For  $i=1, \dots, n$ , let  $L_i$  be the  $n \times n$  diagonal matrix with 1 in every diagonal place except the  $i$ th, which contains  $\lambda_i^2$ , so

$$L_i = \text{diag}(1, 1, \dots, 1, \lambda_i^2, \dots, 1),$$

and  $L_0 = I_n$ . Define the scalars

$$\tau_i = \lambda_i^2 - 1 \quad \text{for } i=1, \dots, n,$$

$$\beta_i = \tau_i / (1 + \tau_i m_{ii}) \quad \text{for } i=1, \dots, n, \quad \text{and}$$

$$\tau_0 = \beta_0 = m_{00} = 0,$$

and the row vectors

$$h_i[1 \times (n-k)] = (p_{i1}, p_{i2}, \dots, p_{i, n-k}) \quad \text{for } i=1, \dots, n, \quad \text{and}$$

$$h_0[1 \times (n-k)] = 0.$$

From this point to the end of Section 4, let  $e_1$  denote a  $p \times 1$  column vector with a one in the first place and zeroes elsewhere,  $e_1 = (1, 0, 0, \dots, 0)'$ . (In Section 2,  $e_1$  was defined to be of this form, but of length  $n$ .)

The covariance relations (4.19) may be expressed more compactly as

$$\text{Cov}(Y^*) = I_p \otimes L_i \quad \text{under } H_i, \quad \text{for } i=1, \dots, n.$$

We may write  $L_i$  as  $I_n + \text{diag}(0, \dots, 0, \tau_i, 0, \dots, 0)$  for  $i=0, 1, \dots, n$ .

It is understood that  $\tau_i$  appears in the  $i$ th diagonal place, except that when  $i=0$ ,  $\tau_0 = 0$  may be said to appear in any position, for instance the first place.

To compute the distribution of  $T$ , we first examine  $Z = P'MY$ , noting that

$$Z^* = (I_p \otimes P'M)Y^*.$$



It is immediate that under  $H_i$ ,  $i=0,1, \dots, n$ ,  $E(Z^*) = 0$  and

$$\begin{aligned} \text{Cov}(Z^*) &= (I_p \otimes P'M)(I_p \otimes L_i)(I_p \otimes P'M)' \\ &= I_p \otimes P'ML_iMP \\ &= I_p \otimes DP'[I_n + \text{diag}(0, \dots, 0, \tau_i, 0, \dots, 0)]PD \\ &= I_p \otimes [D + \tau_i(h_i \ 0)'(h_i \ 0)], \text{ where } 0 \text{ is } O(1 \times k) \end{aligned}$$

The last  $k$  rows of  $Z$  are degenerate at  $O(1 \times p)$ , and the block structure of  $\text{Cov}(Z^*)$  shows that the remaining entries of  $Z$  form  $p$  independent sets of  $n-k$  components. Each of these  $p$  sets is a column of entries  $Z_{1j}, Z_{2j}, \dots, Z_{n-k,j}$ , where  $j$  is fixed between 1 and  $p$ . Therefore, under  $H_i$ , standard normal distribution theory shows that, for every  $j=1,2, \dots, p$ ,

$$\begin{aligned} f_{Z_{1j}, \dots, Z_{n-k,j}}(z) &= (2\pi)^{-(n-k)/2} \left[ \det(I_{n-k} + \tau_i h_i' h_i) \right]^{-\frac{1}{2}} \\ (4.20) \quad &\cdot \exp \left[ -\frac{1}{2} z' (I_{n-k} + \tau_i h_i' h_i)^{-1} z \right] \text{ for all } z[(n-k) \times 1] \end{aligned}$$

To simplify this, note that (see Press, 1972, p. 20)

$$\det(I_{n-k} + \tau_i h_i' h_i) = 1 + \tau_i h_i' h_i = 1 + \tau_i m_{ii} \quad ,$$

the last step following from the fact that  $M = PDP'$ , which shows

$$m_{ii} = \sum_{\ell=1}^{n-k} p_{i\ell}^2 = h_i' h_i \quad \text{for } i=1, \dots, n \quad .$$

Also, by the Woodbury binomial inverse theorem (Press, 1972, p. 23)

$$(I_{n-k} + \tau_i h_i' h_i)^{-1} = I - [\tau_i / (1 + \tau_i m_{ii})] h_i' h_i \quad .$$

So (4.20) can be written as

$$f_{Z_{1j}, \dots, Z_{n-k,j}}(z) = (2\pi)^{-(n-k)/2} [1 + \tau_i m_{ii}]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} z' z + \frac{1}{2} \beta_i z' h_i' h_i z \right] \quad .$$

The distribution of the first  $n-k$  rows of  $Z$  under  $H_1$  is thus the product of  $p$  such distributions. Let  $z^j$  denote the  $(n-k) \times 1$  column vector  $(z_{1j}, \dots, z_{n-k,j})'$ , the  $j$ th column of  $Z$  with the degenerate entries deleted. Then under  $H_1$ ,

$$(4.21) \quad f_{Z_{11}, \dots, Z_{n-k,p}}(z_{11}, \dots, z_{n-k,p}) = (2\pi)^{-(n-k)p/2} (1 + \tau_i m_{ii})^{-p/2} \cdot \exp \left[ -\frac{1}{2} \sum_{j=1}^p z^j{}' z^j + \frac{1}{2} \beta_i \sum_{j=1}^p z^j{}' h_i' h_i z^j \right] .$$

Partitioning the first  $n-k$  rows of  $Z$  into sub-matrices  $Z_1$  and  $Z_2$  as in (4.4), we obtain the distribution of the maximal invariant  $T$  by a routine change of variables. Let  $W^j$  denote the  $j$ th column of  $W$ , so  $z^j = \begin{bmatrix} T \\ I \end{bmatrix} W^j$  for  $j=1, \dots, p$ . Using (4.21) and the Jacobian computed earlier,

$$(4.22) \quad \begin{aligned} f_T(T) &= (2\pi)^{-(n-k)p/2} (1 + \tau_i m_{ii})^{-p/2} \int \exp \left[ -\frac{1}{2} \sum_{j=1}^p W^j{}' (T' \ I) \begin{bmatrix} T \\ I \end{bmatrix} W^j \right. \\ &\quad \left. + \frac{1}{2} \beta_i \sum_{j=1}^p W^j{}' (T' \ I) h_i' h_i \begin{bmatrix} T \\ I \end{bmatrix} W^j \right] |\det W|^{n-k-p} dW \\ &= (2\pi)^{-(n-k)p/2} (1 + \tau_i m_{ii})^{-p/2} \int \exp \left[ -\frac{1}{2} \text{tr } W' Q W \right. \\ &\quad \left. + \frac{1}{2} \beta_i n_i \begin{bmatrix} T \\ I \end{bmatrix} W W' (T' \ I) h_i' h_i \begin{bmatrix} T \\ I \end{bmatrix} \right] |\det W|^{n-k-p} dW . \end{aligned}$$

To simplify the integral, we define the matrix variable  $F(p \times p) = W' Q^{\frac{1}{2}}$  and change variables from  $W$  to  $F$ , finding that

$$f_T(T) = (2\pi)^{-(n-k)p/2} (1 + \tau_i m_{ii})^{-p/2} (\det Q)^{-(n-k)/2} \sigma ,$$

where

$$\sigma = \int \exp \left[ -\frac{1}{2} \text{tr } F F' + \frac{1}{2} \beta_i h_i \begin{bmatrix} T \\ I \end{bmatrix} Q^{-\frac{1}{2}} F' F Q^{-\frac{1}{2}} (T' \ I) h_i' \right] |\det F|^{n-k-p} dF .$$

Defining the row vector  $r_i(1 \times p) = h_i \begin{bmatrix} T \\ I \\ I \\ O \end{bmatrix} Q^{-\frac{1}{2}} = P_i \begin{bmatrix} T \\ I \\ I \\ O \end{bmatrix} Q^{-\frac{1}{2}}$ , we see that

$$\sigma = \int \exp\left[-\frac{1}{2} \text{tr } FF' + \frac{1}{2} \beta_i r_i F' F r_i'\right] |\det F|^{n-k-p} dF .$$

It is a direct consequence of the definition of  $r_i$  that  $r_i r_i' = s_i$ , the scalar defined in (4.14). Therefore, Vinograd's Theorem implies the existence of an orthogonal matrix  $K(p \times p)$  such that

$$r_i = s_i^{\frac{1}{2}} e_1' K ,$$

or equivalently

$$K r_i' = s_i^{\frac{1}{2}} e_1 .$$

Defining  $G(p \times p) = FK'$  and changing variables once more,

$$(4.23) \quad \sigma = \int \exp\left[-\frac{1}{2} \text{tr } GG' + \frac{1}{2} \beta_i s_i e_1' G' G e_1\right] |\det G|^{n-k-p} dG .$$

It is clear that for fixed  $n$ ,  $k$ , and  $p$ ,  $\sigma$  is a function of the single scalar argument  $\beta_i s_i$ . We make this explicit by writing  $\sigma(\beta_i s_i)$  for  $\sigma$ .

Theorem 4.3: For the multivariate regression model, variance slippage single-outlier problem, the distribution of  $T$  under  $H_i$ ,  $i=0,1, \dots, n$ , is

$$(4.24) \quad f_T(T) = (2\pi)^{-(n-k)p/2} (1 + \tau_i m_{ii})^{-p/2} (\det Q)^{-(n-k)/2} \sigma(\beta_i s_i) ,$$

where  $s_i$  is defined by (4.14), and  $\sigma(\beta_i s_i)$  by (4.23).

## 5. Studentized Residual Rules for the Single-Outlier Problem

We define squared studentized residuals to be the diagonal entries of the  $n \times n$  matrix  $R(R'R)^{-1}R'$ , normalized by the constants  $m_{ii}^{-1}$ . It will be shown that simple rules based on the squared studentized residuals are invariant admissible for both the mean slippage single-outlier problem and the variance slippage single-outlier problem.

Definition 5.1: For  $i=1,2, \dots, n$ , the  $i$ th squared studentized residual is

$$(5.1) \quad v_i^2 = m_{ii}^{-1} R_i S^{-1} R_i' ,$$

where  $m_{ii}$  is the  $i$ th diagonal entry of  $M$ ,  $R_i (1 \times p)$  is the  $i$ th row of  $R$ , and  $S(p \times p) = R'R$ .

Lemma 5.1: The scalar  $s_i$  defined by (4.14) equals  $m_{ii} v_i^2$  for  $i=1, \dots, n$ .

Proof: This follows immediately from  $R=PZ$  and (4.10). QED

Definition 5.2: For the multivariate regression model, mean slippage single-outlier problem, let  $f_1, \dots, f_n$  be arbitrary positive constants. Let  $\mathfrak{U}$  denote the decision rule which states: take action  $D_0$  when

$$\max_j f_j m_{jj} v_j^2 < K ;$$

for  $i=1, \dots, n$ , take action  $D_i$  when

$$(5.2) \quad f_i m_{ii} v_i^2 = \max_j f_j m_{jj} v_j^2 \geq K .$$

The constant  $K$  is determined by the desired value of  $\Pr(D_0|H_0)$ . If more than one value of  $i$  satisfies (5.2), the choice among the corresponding  $D_i$  may be made in any prespecified fashion, possibly at random.

The distribution of  $T$  under  $H_i$  will be denoted by  $f_T^i(T)$ . We now show by a method due to Karlin and Truax (1960, Sec. 9) that the integral  $\rho(s_i, \delta_i)$  is a monotone increasing function of  $s_i$  for any positive  $\delta_i$ .

Lemma 5.2: The integral  $\rho(s_i, \delta_i)$  of (4.18) is strictly increasing in  $s_i$  for each positive value of  $\delta_i$ , and thus is a strictly increasing function of the single argument  $s_i \delta_i$ .

Proof: Expanding the integrand in a Taylor series gives

$$\exp\left[-\frac{1}{2}\text{tr}(GG') + g_{11}s_i^{\frac{1}{2}}\delta_i^{\frac{1}{2}}\right] = \exp\left(-\frac{1}{2}\sum_{ij}g_{ij}^2\right)\left[1 + g_{11}s_i^{\frac{1}{2}}\delta_i^{\frac{1}{2}} + \frac{1}{2}g_{11}^2s_i\delta_i + \dots\right].$$

Integrating term by term,

$$\rho(s_i, \delta_i) = \sum_{n=0}^{\infty} C_n s_i^{n/2} \delta_i^{n/2},$$

where

$$C_n = \frac{1}{n!} \int \exp\left(-\frac{1}{2}\sum_{ij}g_{ij}^2\right) g_{11}^n |\det G|^{n-k-p} dG.$$

It is clear that  $C_{2n} \geq 0$  for any  $n \geq 0$ , for the integrand is non-negative. To show that  $C_{2n+1} = 0$  for all  $n$ , change variables from  $G$  to  $H(p \times p) = \text{diag}(-1, 1, 1, \dots, 1)G$ . Then

$$\begin{aligned} C_{2n+1} &= \frac{1}{(2n+1)!} \int \exp\left(-\frac{1}{2}\sum_{ij}g_{ij}^2\right) g_{11}^{2n+1} |\det G|^{n-k-p} dG \\ &= \frac{1}{(2n+1)!} \int \exp\left(-\frac{1}{2}\sum_{ij}h_{ij}^2\right) (-h_{11}^{2n+1}) |\det H|^{n-k-p} dH = -C_{2n+1}. \end{aligned}$$

This establishes that  $C_{2n+1} = 0$ , for all  $n$ , and therefore that

$$\rho(s_i, \delta_i) = \sum_{n=0}^{\infty} C_{2n} s_i^n \delta_i^n, \quad \text{where } C_{2n} \geq 0 \quad \text{for all } n.$$

Hence  $\rho$  is strictly increasing in  $x_i$  for any positive  $\delta_i$ , and also strictly increasing in  $s_i \delta_i$ . QED

This lemma allows us to prove the following:

Theorem 5.1: For the multivariate regression model, mean slippage single-outlier problem, let  $\delta_i = a_i \Sigma^{-1} a_i'$  be known up to a constant for  $i=1, 2, \dots, n$ . That is,  $f_i = \delta_i / \sum_{k=1}^n \delta_k$  is known for  $i=1, \dots, n$ . Then the decision rule  $\bar{u}$  is invariant admissible for the outlier problem for any set of  $\delta_i$ 's proportional to the given  $f_i$ 's. In fact, it is Bayes with respect to any prior distribution in which the  $p_i$ 's are proportional to  $\alpha_i^{-1}$  for  $i=1, \dots, n$ , and all  $p_i$ 's are non-zero. The value of  $i$  satisfying (5.2) is unique with probability one.

Proof: Assume that  $B$ ,  $\Sigma$ , and  $a_1, \dots, a_n$  are known, so we are dealing with simple hypotheses throughout. It will be seen that these parameters enter the Bayes decision rule in the statement of the theorem only through  $\delta_i$ ,  $i=1, \dots, n$ . Let a prior distribution be defined by assigning non-zero probability  $p_j$  to  $H_j$  for  $j=0, 1, \dots, n$ , so that  $\sum_{j=0}^n p_j = 1$ . Any rule  $\varphi$  for which  $\varphi_i(T) = 0$  for all  $i$  such that  $p_i f_T^i(T) < \max_j p_j f_T^j(T)$  is Bayes with respect to the prior  $(p_0, p_1, \dots, p_n)$ . There is essentially only one such rule, since the index  $i$  such that  $p_i f_T^i(T) = \max_j p_j f_T^j(T)$  is unique with probability one. On the set of measure zero where this uniqueness does not hold, any randomization among those  $i$ 's maximizing  $p_i f_T^i(T)$  gives a Bayes rule.

Let  $\omega_i$  denote the region of the sample space on which action  $D_i$  is taken. The Bayes rule with respect to  $(p_0, \dots, p_n)$  is specified, using the distribution  $f_T^i(T)$  of (4.17), by

$$(5.3) \quad \omega_0 = \bigcap_{j=1}^n \{p_0 \rho(0,0) > p_j \alpha_j \rho(s_j, \delta_j)\} \quad ,$$

$$\omega_i = \{p_i \alpha_i \rho(s_i, \delta_i) > p_0 \rho(0,0)\} \cap \bigcap_{\substack{j=1 \\ j \neq i}}^n \{p_i \alpha_i \rho(s_i, \delta_i) > p_j \alpha_j \rho(s_j, \delta_j)\}$$

for  $i=1, \dots, n$  .

Any rule of this form, being Bayes with respect to a given prior, is invariant admissible.

Now consider a prior distribution satisfying

$$(5.4) \quad p_j = c_0 \alpha_j^{-1} = c_0 \exp[\frac{1}{2} m_{jj} \delta_j] \quad \text{for } j=1, \dots, n \quad ,$$

$$p_0 = 1 - \sum_{j=1}^n p_j \quad ,$$

where the constant  $c_0$  is chosen to make all of these positive. For this prior, the region  $\omega_0$  of (5.3) simplifies to

$$\begin{aligned} \omega_0 &= \{p_0 \rho(0,0) > c_0 \rho(s_j, \delta_j) \text{ for } j=1, \dots, n\} \\ &= \{\rho(s_j, \delta_j) < c_0^{-1} p_0 \rho(0,0) \text{ for } j=1, \dots, n\} \\ &= \{\max_j \rho(s_j, \delta_j) < K'\}, \quad \text{where } K' = c_0^{-1} p_0 \rho(0,0) \quad . \end{aligned}$$

The constant  $K'$  depends only on the prior (5.4) and on  $n-k$  and  $p$  through  $\rho(0,0)$  . Viewing  $\rho$  as a strictly increasing function  $g$  of the single argument  $s_j \delta_j$  as in Lemma 5.2,

$$\begin{aligned} \omega_0 &= \{\max_j g(s_j \delta_j) < K'\} \\ &= \{\max_j s_j \delta_j < K''\}, \quad \text{where } K'' = g^{-1}(K') \quad . \end{aligned}$$

And applying Lemma 5.1 gives

$$(5.5) \quad \omega_0 = \{\max_j \delta_j m_{jj} V_j^2 < K''\} = \{\max_j f_j m_{jj} V_j^2 < K\}, \quad \text{where } K = K'' / \sum_k \delta_k \quad .$$

Similar calculations can be made for  $\omega_i$ ,  $i=1, \dots, n$ , giving

$$\begin{aligned}
 \omega_i &= \{p_i \alpha_i \rho(s_i, \delta_i) > p_{00}(0,0), \text{ and } p_i \alpha_i \rho(s_i, \delta_i) > p_j \alpha_j \rho(s_j, \delta_j) \\
 &\quad \text{for all } j \neq 0, i\} \\
 &= \{c_0 \rho(s_i, \delta_i) > p_{00}(0,0), \text{ and } \rho(s_i, \delta_i) > \rho(s_j, \delta_j) \text{ for all } j \neq 0, i\} \\
 &= \{\rho(s_i, \delta_i) = \max_{j \neq 0} \rho(s_j, \delta_j) > c_0^{-1} p_{00}(0,0) = K'\} \\
 &= \{s_i \delta_i = \max_j s_j \delta_j > K'' = g^{-1}(K')\} \\
 &= \{\delta_i m_{ii} V_i^2 = \max_j \delta_j m_{jj} V_j^2 > K''\} \quad ;
 \end{aligned}$$

$$(5.6) \quad \omega_i = \{f_i m_{ii} V_i^2 = \max_j f_j m_{jj} V_j^2 > K = K'' / \sum_k \delta_k\} \quad .$$

Since the  $f_i$ 's are known,  $f_i m_{ii} V_i^2$  can be computed for every  $i = 1, \dots, n$ , and (5.5) and (5.6) specify a decision rule. The form of this rule does not depend on  $\sum_k \delta_k$ , so it is invariant admissible for all values of  $\sum_k \delta_k$ . QED

The dependence of the various expressions in the proof of Theorem 5.1 is clarified by Figure 5.1, in which functional dependence is indicated by an arrow:

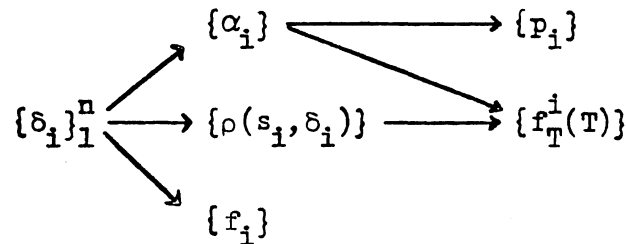


Figure 5.1. Dependence of Expressions in the Proof of Theorem 5.1.

This illustrates the fact that any family  $\{f_i\}_{i=1}^n$  corresponds to an infinitude of families  $\{\delta_i\}_1^n$ , each of which leads to a different  $\{f_T^i(T)\}_1^n$  and a different collection of priors  $\{p_i\}_1^n$ . Only  $\{f_i\}_1^n$  must be known to perform the test of Theorem 5.1. As noted earlier, the parameters  $B$ ,  $\Sigma$ , and  $a_i$ ,  $i = 1, \dots, n$  of the mean slippage single-outlier problem enter the discussion only through the  $n$  scalar quantities  $\delta_i$ ,  $i = 1, \dots, n$ .



Two particular cases are of special interest.

Corollary 5.1: The decision rule: take action  $D_0$  when  $\max_j V_j^2 < K$ ; take action  $D_i$  when  $V_i^2 = \max_j V_j^2 \geq K$ , is invariant admissible for the multivariate regression model, mean slippage single-outlier problem with  $\delta_i$  proportional to  $m_{ii}^{-1}$  for  $i=1, \dots, n$ . Furthermore, this rule is Bayes with respect to any prior distribution which assigns equal non-zero probabilities to  $H_1, \dots, H_n$ , as well as non-zero probability to  $H_0$ .

Corollary 5.2: The decision rule: take action  $D_0$  when  $\max_j m_{jj} V_j^2 < K$ ; take action  $D_i$  when  $m_{ii} V_i^2 = \max_j m_{jj} V_j^2 \geq K$ , is invariant admissible for the multivariate regression model, mean slippage single-outlier problem with all  $\delta_i$  equal.

These are proved by routine substitution into Theorem 5.1. An interesting feature of Corollary 5.1 is that whenever  $\delta_i$  varies directly as  $m_{ii}^{-1}$  for  $i=1, \dots, n$ , no matter what the constant of proportionality, (5.4) gives the same family of prior distributions, indexed only by a single constant  $c_1$ :

$$p_j = c_1 \quad \text{for } j=1, \dots, n,$$

$$p_0 = 1 - nc_1, \quad \text{where } 0 < c_1 < 1/n.$$

The scalar  $\delta_i = a_i \sum a_i^{-1}$  indicates the distance by which the ith observation departs from the null hypothesis model under  $H_i$ . Theorem 5.1 shows that we can generate a large class of invariant admissible rules by changing the sizes of these  $n$  distances relative to each other. It is interesting and a bit surprising that we can show all of these rules to be invariant admissible for the single-outlier problem with a single, fixed set of scalars  $\delta_1, \dots, \delta_n$ . This is established in

Theorem 5.2: For the mean slippage single-outlier problem, fix  $\delta_i = a_i \Sigma^{-1} a_i'$  for  $i=1, \dots, n$ . Then for any choice of  $f_1, \dots, f_n$ , the decision rule  $\bar{U}$  is invariant admissible for this outlier problem. In fact, it is Bayes with respect to any prior distribution of the type given in (5.7). The value of  $i$  satisfying (5.2) is unique with probability one.

Proof: We follow the proof of Theorem 5.1, but now find the regions obtained from (5.3) by considering a prior distribution satisfying

$$(5.7) \quad p_j = c_0 \alpha_j^{-1} \rho(s_j, f_j) / \rho(s_j, \delta_j) \quad \text{for } j=1, \dots, n, \\ p_0 = 1 - \sum_{j=1}^n p_j, \quad \text{where} \\ 0 < c_0 < 1 / \sum_{j=1}^n \left[ \alpha_j^{-1} \rho(s_j, f_j) / \rho(s_j, \delta_j) \right].$$

(The integral  $\rho(s_j, f_j)$  in  $p_j$  could be replaced by the more general  $\rho(s_j, c_1 f_j)$ , where  $c_1$  is a positive constant which does not vary with  $j$ .) The region  $\omega_0$  of (5.3) then simplifies to

$$\omega_0 = \{p_0 \rho(0,0) > c_0 \rho(s_j, f_j) \text{ for } j=1, \dots, n\},$$

and by the analysis of  $\omega_0$  in Theorem 5.1 with  $\delta_j$  replaced by  $f_j$ ,

$$\omega_0 = \{\max_j s_j f_j < K\} = \{\max_j f_j m_{jj} V_j^2 < K\}.$$

Similarly, for  $i=1, \dots, n$ , the region  $\omega_i$  of (5.3) can be evaluated under the prior distribution of (5.7) as

$$\omega_i = \{c_0 \rho(s_i, f_i) > p_0 \rho(0,0), \text{ and } \rho(s_i, f_i) > \rho(s_j, f_j) \\ \text{for all } j \neq 0, i\} \\ = \{f_i m_{ii} V_i^2 = \max_j f_j m_{jj} V_j^2 \geq K\} \quad \text{QED}$$

It will now be shown that the rule of Theorem 5.1 is invariant admissible for the variance slippage single-outlier problem. We begin by establishing the monotonicity of  $\sigma(\beta_i s_i)$ .

Lemma 5.3: The integral  $\sigma(\beta_i s_i)$  of (4.24) is a strictly increasing function of the scalar argument  $\beta_i s_i$ .

Proof: The non-negative integrand depends on the argument only through the term  $\exp(\frac{1}{2}\beta_i s_i e_1' G e_1)$ . This term is strictly increasing in  $\beta_i s_i$  when  $g_{i1} \neq 0$  for some  $i=1, \dots, p$ , since  $e_1' G e_1 = \sum_{i=1}^p g_{i1}^2$ . QED

Theorem 5.3: For the multivariate regression model, variance slippage single-outlier problem, let  $\beta_i$  of Definition 4.4 be known up to a constant for  $i=1, 2, \dots, n$ . That is,  $f_i = \beta_i / \sum_{i=1}^n \beta_i$  is known for  $i=1, \dots, n$ . Then the decision rule  $\tilde{U}$  is invariant admissible for the outlier problem for any set of  $\beta_i$ 's proportional to the given  $f_i$ 's. In fact, it is Bayes with respect to any prior distribution of the type (5.9). The value of  $i$  satisfying (5.2) is unique with probability one.

Proof: The method used to prove Theorem 5.1 is applicable here, so the proof will only be outlined, with emphasis on the computations, which differ from those done earlier. The Bayes rule with respect to the prior distribution  $(p_0, p_1, \dots, p_n)$  is specified by

$$\omega_i = \bigcap_{j \neq i} \{p_i f_T^i(T) > p_j f_T^j(T)\}.$$

Using the distribution of  $T$  under  $H_i$ ,  $i=0, 1, \dots, n$ , from (4.24), we obtain

$$\begin{aligned} \omega_0 &= \{p_0 \sigma(0) > p_j (1 + \tau_j m_{jj})^{-p/2} \sigma(\beta_j s_j) \text{ for } j=1, \dots, n\}, \\ \omega_i &= \{p_i (1 + \tau_i m_{ii})^{-p/2} \sigma(\beta_i s_i) > p_0 \sigma(0)\} \cap \bigcap_{\substack{j=1 \\ j \neq i}}^n \{p_i (1 + \tau_i m_{ii})^{-p/2} \sigma(\beta_i s_i) \\ &\quad > p_j (1 + \tau_j m_{jj})^{-p/2} \sigma(\beta_j s_j)\}. \end{aligned} \quad (5.8)$$

Choose a prior distribution of the form

$$(5.9) \quad \begin{aligned} p_j &= c_0(1 + \tau_j m_{jj})^{p/2} \quad \text{for } j=1, \dots, n, \\ p_0 &= 1 - \sum_{j=1}^n p_j, \quad \text{where} \\ 0 &< c_0 < 1 / \sum_{j=1}^n (1 + \tau_j m_{jj})^{p/2}. \end{aligned}$$

All  $n+1$  elements of this prior distribution are positive. The regions of (5.8) become

$$(5.10) \quad \begin{aligned} \omega_0 &= \{p_0 \sigma(0) > c_0 \sigma(\beta_j s_j) \text{ for } j=1, \dots, n\} \\ &= \{ \max_{1 \leq j \leq n} \sigma(\beta_j s_j) < K' \} \quad \text{where } K' = c_0^{-1} p_0 \sigma(0) \\ &= \{ \max_j \beta_j m_{jj} V_j^2 < \sigma^{-1}(K') \} \\ &= \{ \max_j f_j m_{jj} V_j^2 < K \} \quad \text{where } K = \sigma^{-1}(K') / \sum_{j=1}^n \beta_j ; \\ \omega_1 &= \{c_0 \sigma(\beta_1 s_1) > p_0 \sigma(0), \text{ and } c_0 \sigma(\beta_1 s_1) > c_0 \sigma(\beta_j s_j) \text{ for all } j \neq 0, 1\} \\ &= \{ \sigma(\beta_1 s_1) = \max_{1 \leq j \leq n} \sigma(\beta_j s_j) > c_0^{-1} p_0 \sigma(0) \} \\ &= \{ \beta_1 s_1 = \max_j \beta_j s_j > \sigma^{-1}(K') \} \\ &= \{ f_{11} m_{11} V_1^2 = \max_j f_j m_{jj} V_j^2 > K = \sigma^{-1}(K') / \sum_{j=1}^n \beta_j \} . \end{aligned}$$

This decision rule is Bayes with respect to any prior distribution of the type (5.9), and therefore is invariant admissible no matter what the value of  $\sum_{j=1}^n \beta_j$ . QED

The rules of Corollaries 5.1 and 5.2 are invariant admissible for the variance slippage problem, as substitution into Theorem 5.3 shows. These results could be stated as corollaries, with  $\delta_i$  replaced by  $\beta_i$ .

Theorem 5.3 shows that we can generate a class of invariant admissible decision rules by changing the sizes of the  $n$  departures from

hypothesis  $H_0$  of the variance slippage single-outlier problem. The following theorem verifies that the entire class of rules from Theorem 5.3 is invariant admissible under the variance slippage single-outlier problem with a single fixed set of  $n$  constants  $\lambda_1^2, \dots, \lambda_n^2$ .

Theorem 5.4: For the multivariate regression model, variance slippage single-outlier problem, fix  $\lambda_i^2$  for  $i=1, \dots, n$ . Let  $f_1, \dots, f_n$  be arbitrary positive constants. Then the decision rule  $\bar{3}$  is invariant admissible for this outlier problem. In fact, it is Bayes with respect to any prior distribution as in (5.11). The value of  $i$  satisfying (5.2) is unique with probability one.

Proof: Fixing  $\lambda_i^2$ ,  $i=1, \dots, n$  is equivalent to fixing either  $\tau_i$ ,  $i=1, \dots, n$ , or  $\beta_i$ ,  $i=1, \dots, n$  where these are as in Definition 4.4. We proceed to find the regions obtained from (5.8) after choosing a prior distribution of the form

$$(5.11) \quad p_j = c_0(1 + \tau_j m_{jj})^{p/2} \sigma(f_j s_j) / \sigma(\beta_j s_j) \quad \text{for } j=1, \dots, n, \\ p_0 = 1 - \sum_{j=1}^n p_j,$$

where  $c_0$  is selected to make all  $n+1$  components of the prior distribution positive. (The integral  $\sigma(f_j s_j)$  in  $p_j$  could be replaced by the more general term  $\sigma(c_1 f_j s_j)$ , where  $c_1$  is any positive constant.) Substituting the expressions from (5.11) into (5.8) and noting the resemblance of the resulting terms to (5.10), we easily obtain

$$\omega_0 = \{p_0 \sigma(0) > c_0 \sigma(f_j s_j) \text{ for } j=1, \dots, n\} \\ = \{\max_j f_j m_{jj} V_j^2 < K\},$$

and

$$\omega_i = \{c_0 \sigma(f_i s_i) > p_0 \sigma(0), \text{ and } c_0 \sigma(f_i s_i) > c_0 \sigma(f_j s_j) \text{ for all } j \neq 0, i\} \\ = \{f_i m_{ii} V_i^2 = \max_j f_j m_{jj} V_j^2 > K\}. \quad \text{QED}$$

## 6. Studentized Rules for Mean Slippage with Several Outliers in the Same Direction

The model of Definition 2.3, with equation  $Y = XB + A + U$ , is considered in this section. A class of optimal rules for testing alternatives involving several outliers is developed here, using the methods of Sections 4 and 5. These rules are of theoretical interest, although

the assumptions made in specifying the alternatives may be unrealistic from a practical point of view. We begin with a brief discussion of these assumptions.

It is supposed that there are either  $h$  outliers, where  $h$  is a known number less than  $n/2$ , or none. If there are  $h$  outliers, they have all slipped in a common direction. In terms of the model, all non-zero rows of  $A$  must be scalar multiples of a common row vector  $a(1 \times p)$ . Let these  $h$  non-zero scalars be denoted by  $c_1, \dots, c_h$ , where  $c_1 \leq c_2 \leq \dots \leq c_h$ , so the non-zero rows of  $A$  are  $c_i a$ ,  $i = 1, \dots, h$ . It is assumed that the  $c_i$ 's are known.

Definition 6.1: Let  $v = [v(1), v(2), \dots, v(n)]$  be a permutation of the first  $n$  positive integers. The permuted identity matrix corresponding to  $v$ , denoted  $I^v$ , is an  $n \times n$  matrix whose  $(j, k)$ th element is  $\delta_{v(j), k}$ , where  $\delta$  is the Kronecker delta. That is,

$$I^v(j, k) = \begin{cases} 1 & \text{if } k = v(j) \\ 0 & \text{otherwise} \end{cases}.$$

The hypothesis of no outliers can be expressed as  $A = 0$ . Any hypothesis of  $h$  outliers can be written in the form

$$(6.1) \quad A = I^v(c_1 \ c_2 \ \dots \ c_h \ 0 \ 0 \ \dots \ 0)'a \ ,$$

where  $I^v$  corresponds to some permutation  $v$ . Pre-multiplying any  $n \times 1$  column vector by  $I^v$  moves its  $v(i)$ th entry to the  $i$ th place for every  $i$ . Thus the permutation  $v$  specifies which  $h$  observations correspond to the respective slippage constants  $c_i$ .

The unidirectional  $h$ -outlier problem for the multivariate regression model with mean slippage can be formally specified as the single-outlier problem was in Definition 2.7. The hypotheses are  $H_0$ , that  $A = 0$ , and  $H_{i_1, \dots, i_h}$ , that rows  $i_1, \dots, i_h$  of  $A$  are  $c_1 a, \dots, c_h a$ , respectively, all other rows being zero. The elements of the action space correspond to these hypotheses. The state space consists of elements  $\theta = (i_1, \dots, i_h, c_1, \dots, c_h, a, B, \Sigma)$ , and the loss function is zero-one. Filling in the details is an exercise, and will be omitted.

The set of all permutations of the set  $\{1, 2, \dots, n\}$  can be partitioned into classes in such a way that two permutations  $u$  and  $v$  are in the same class if and only if

$$I^u(c_1 \dots c_h 0 \dots 0)' = I^v(c_1 \dots c_h 0 \dots 0)' \quad .$$

This results in an equivalence relation, each class of which corresponds to a different matrix of the type (6.1). Each  $h$ -outlier alternative  $H_{i_1, \dots, i_h}$  of the problem is therefore associated with a class of permutations. For notational convenience, we will instead associate each of these alternatives with a single permutation  $v$  in its class, and refer to it as  $H_v$ . The number of distinct alternatives of this type is not important in the ensuing analysis, and will not be considered further.

The distribution of the maximal invariant  $T$  will now be found under the hypothesis  $H_v$ , and then used to derive a class of decision rules which are Bayes with respect to specified prior distributions, and there-

fore invariant admissible. Our approach will be to specialize the results of Section 4 for general A to the h-outlier case rather than to the single-outlier case.

**Lemma 6.1:** In the multivariate regression model with mean slippage, let  $z_i$ ,  $a_i$ , and  $R_i$  denote the  $i$ th rows of Z, A, and R, and  $P^j$  and  $P_i$  the  $j$ th column and  $i$ th row of P. Then

$$\sum_{j=1}^{n-k} z_j \Sigma^{-1} z_j' = \text{tr}[W \Sigma^{-1} W' Q] \quad ,$$

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} z_j' = \text{tr}[A \Sigma^{-1} W' (T' \ I \ 0) P'] = \sum_{i=1}^n a_i \Sigma^{-1} W' (T' \ I \ 0) P_i' \quad ,$$

and

$$\sum_{j=1}^{n-k} P^j A \Sigma^{-1} A' P^j = \text{tr}[M A \Sigma^{-1} A'] \quad .$$

**Proof:** Contained in the proof of Lemma 4.1.

From (4.9), (4.10), (4.11), and Lemma 6.1, the distribution of T is

$$f_T(T) = (2\pi)^{-(n-k)p/2} (\det \Sigma)^{-(n-k)/2} \text{etr}[-\frac{1}{2} M A \Sigma^{-1} A'] \\ \cdot \int \exp[-\frac{1}{2} \text{tr}(W \Sigma^{-1} W' Q) + \text{tr}(A \Sigma^{-1} W' (T' \ I \ 0) P')] | \det W |^{n-k-p} dW \quad .$$

Change variables from W to  $E(p \times p) = \Sigma^{-\frac{1}{2}} W' Q^{\frac{1}{2}}$ , and define

$$\tau_i(p \times 1) = \Sigma^{-\frac{1}{2}} a_i' \quad \text{for } i=1, \dots, n \quad .$$

Substitution shows the distribution of T under the model  $Y = XB + A + U$  to be

$$(6.2) \quad f_T(T) = (2\pi)^{-(n-k)p/2} (\det Q)^{-(n-k)/2} \text{etr}(-\frac{1}{2} M A \Sigma^{-1} A') \rho \quad ,$$

where

$$\rho = \int \text{etr} \left[ -\frac{1}{2} E E' + E Q^{-\frac{1}{2}} (T' \ I \ 0) \sum_{i=1}^n P_i' \tau_i' \right] | \det E |^{n-k-p} dE \quad .$$



This formula, which holds for all A, is identical to (4.13) except that  $P_i' \tau'$  has been replaced by  $\sum_{i=1}^n P_i' \tau'_i$ .

We now consider the form of (6.2) under an alternative of the type found in the h-outlier problem. The alternative  $H_v$ , as developed so far, specifies that A can be written as in (6.1). We now generalize the form of the matrix A by allowing it to be multiplied by a constant  $\delta_v^{\frac{1}{2}}$ , which varies from one h-outlier alternative to another. Then  $H_v$  becomes

$$(6.3) \quad A = \delta_v^{\frac{1}{2}} I^v (c_1 \dots c_h 0 \dots 0)' a \quad .$$

With  $a(1 \times p)$  as in (6.3), define  $\tau(p \times 1) = \Sigma^{-\frac{1}{2}} a'$ . Also, define scalars  $c_i = 0$  for  $i = h+1, \dots, n$ , so  $A = \delta_v^{\frac{1}{2}} I^v (c_1 \dots c_n)' a$ . Then

$$\tau_i = \Sigma^{-\frac{1}{2}} a'_i = \delta_v^{\frac{1}{2}} c_{v(i)} \Sigma^{-\frac{1}{2}} a' = \delta_v^{\frac{1}{2}} c_{v(i)} \tau \quad ,$$

and

$$\sum_{i=1}^n P_i' \tau'_i = \delta_v^{\frac{1}{2}} \left[ \sum_{i=1}^n c_{v(i)} P_i' \right] \tau' \quad .$$

This is the product of non-null matrices of size  $n \times 1$  and  $1 \times p$ , and therefore has rank one. Consequently, the steps in the one-outlier case of Section 4 can be paralleled here. Define the matrix

$$(6.4) \quad N(p \times p) = \left[ \sum_{i=1}^n \tau_i P_i' \right] \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} Q^{-1} (T' \ I \ 0) \left[ \sum_{i=1}^n P_i' \tau'_i \right] = \delta_v s_v \tau \tau'$$

where the scalar

$$s_v = \left[ \sum_{i=1}^n c_{v(i)} P_i' \right] \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} Q^{-1} (T' \ I \ 0) \left[ \sum_{i=1}^n c_{v(i)} P_i' \right] \quad .$$

N is clearly of rank one, and the method of Section 4 shows that  $\rho$  is a function of the scalar arguments  $s_v$  and  $\delta_v \tau' \tau$ :

$$(6.5) \quad \rho(s_v, \delta_v \tau' \tau) = \int \exp \left[ -\frac{1}{2} \text{tr}(GG') + g_{11} s_v^{\frac{1}{2}} (\delta_v \tau' \tau)^{\frac{1}{2}} \right] |\det G|^{n-k-p} dG \quad .$$

To obtain the distribution  $f_T^V(T)$  of  $T$  under  $H_V$ , it remains only to note that

$$\begin{aligned} \text{tr}[MA\Sigma^{-1}A'] &= \delta_V \text{tr}[MI^V(c_1 \dots c_n)'a\Sigma^{-1}a'(c_1 \dots c_n)I^V] \\ &= \delta_V \tau' \tau [c_{V(1)} \dots c_{V(n)}] M [c_{V(1)} \dots c_{V(n)}]' \end{aligned}$$

Combining these results, we have

$$(6.6) \quad f_T^V(T) = (2\pi)^{-(n-k)p/2} (\det Q)^{-(n-k)/2} \alpha_V \rho(s_V, \delta_V \tau' \tau),$$

where

$$(6.7) \quad \alpha_V = \exp\left[-\frac{1}{2} \delta_V \tau' \tau [c_{V(1)} \dots c_{V(n)}] M [c_{V(1)} \dots c_{V(n)}]'\right]$$

**Lemma 5.2** shows that  $\rho$  is an increasing function of  $s_V$  for any positive value of  $\delta_V \tau' \tau$ , and is a strictly increasing function of the single argument  $\delta_V s_V \tau' \tau$ . The next lemma provides a more convenient expression for  $s_V$ .

**Lemma 6.2:** The scalar  $s_V$  defined by (6.4) is given under  $H_V$  by

$$(6.8) \quad s_V = \left[ \sum_{i=1}^n c_{V(i)} R_i \right] (R'R)^{-1} \left[ \sum_{i=1}^n c_{V(i)} R_i \right]'$$

**Proof:** This is a direct result of the definition of  $s_V$  and the consequences of (4.10) mentioned in the proof of Lemma 5.1. QED

A decision rule which is Bayes, and therefore invariant admissible, for the  $h$ -outlier problem will now be derived. Recall that the action space of the  $h$ -outlier decision problem consists of elements  $D_V$ : act as if  $H_V$  is true, one for each hypothesis  $H_V$ , and  $D_0$ : act as if  $H_0$  is true.

Theorem 6.1: In the multivariate regression model with mean slippage, take  $h < n/2$ , and assume that the scalars  $c_1, \dots, c_h$  are non-zero. Define  $c_i = 0$  for  $i = h+1, \dots, n$ . The alternative hypotheses are  $H_0: A = 0$ , and  $H_v: A = \frac{1}{\sigma_v^2} I^v (c_1 \dots c_n)' a$ , for any set of permutations  $v$  yielding all distinct matrices  $A$  of the form (6.3). Hence  $H_0$  is the hypothesis of no outliers, and  $H_v$  is the  $h$ -outlier hypothesis that  $A$  has rows

$$a_i = \frac{1}{\sigma_v^2} c_{v(i)} a \quad .$$

Let  $\{f_v\}$  be an arbitrary set of positive constants corresponding to the hypotheses  $H_v$ , and consider the decision rule which states: take action  $D_0$  when

$$\max_u f_u s_u < K \quad ;$$

for each  $v$ , take action  $D_v$  when

$$(6.9) \quad f_v s_v = \max_u f_u s_u \geq K \quad .$$

The constant  $K$  is determined by the desired value of  $\Pr(D_0|H_0)$ . If more than one  $v$  satisfies (6.9), which happens with probability zero, the choice among the corresponding  $D_v$  may be made in any pre-specified fashion, possibly at random. Any decision rule of this form is invariant admissible for this outlier problem. In fact, it is Bayes with respect to any prior distribution of the type given in (6.10).

Proof: The general method used to prove Theorem 5.1 is applicable here. Assume that the parameters  $B$ ,  $\Sigma$ , and  $a$  are known, so we are dealing with simple hypotheses throughout. Let a prior distribution be defined by assigning non-zero probabilities  $p_0$  to  $H_0$  and  $p_v$  to each  $H_v$ . The permutation  $v$  such that  $p_v f_T^v(T) = \max_u p_u f_T^u(T)$  is unique with probability

one (considering the situation of  $H_0$  as a permutation for purposes of this statement), so there is essentially only one Bayes decision rule with respect to the prior distribution determined by  $p_0$  and the set of  $p_v$ 's .

Let  $\omega_v$  denote the region of the sample space on which action  $D_v$  is taken. Using the distribution  $f_T^v(T)$  of (6.6), we find that the Bayes rule with respect to the prior distribution is

$$\omega_0 = \bigcap_u \{p_0 \rho(0,0) > p_u \alpha_u \rho(s_u, \delta_u \tau' \tau)\} ,$$

$$\omega_v = \{p_v \alpha_v \rho(s_v, \delta_v \tau' \tau) > p_0 \rho(0,0)\} \cap \bigcap_{u \neq v} \{p_v \alpha_v \rho(s_v, \delta_v \tau' \tau) > p_u \alpha_u \rho(s_u, \delta_u \tau' \tau)\} \\ \text{for every } v .$$

Any rule of this form, being Bayes with respect to a given prior distribution, is invariant admissible.

Now choose a prior distribution

$$(6.10) \quad p_v = c_0 \alpha_v^{-1} \rho(s_v, f_v) / \rho(s_v, \delta_v \tau' \tau) \quad \text{and} \quad p_0 = 1 - \sum_v p_v ,$$

where  $\alpha_v$  is defined by (6.7), and the constant  $c_0$  is chosen to make all of these positive. Then

$$\begin{aligned} \omega_0 &= \{p_0 \rho(0,0) > c_0 \rho(s_u, f_u) \text{ for all } u\} \\ &= \{\max_u \rho(s_u, f_u) < K' = c_0^{-1} p_0 \rho(0,0)\} \\ &= \{\max_u f_u s_u < K\} . \end{aligned}$$

And for each permutation  $v$ ,

$$\begin{aligned} \omega_v &= \{c_0 \rho(s_v, f_v) > p_0 \rho(0,0), \text{ and } \rho(s_v, f_v) > \rho(s_u, f_u) \text{ for all } u \neq v\} \\ &= \{\rho(s_v, f_v) = \max_u \rho(s_u, f_u) > K'\} \\ &= \{f_v s_v = \max_u f_u s_u > K\} . \end{aligned}$$

QED

Further generalizations of this theorem are easily obtained. First, since the row vector  $a(1 \times p)$  entered the distribution of  $T$  and the proof of Theorem 6.1 only through the scalar  $\tau'\tau$ , the discussion applies without change if  $a(1 \times p)$  is replaced in  $H_v$  by  $a_v(1 \times p)$ , which changes with  $v$ , as long as the scalar  $a_v \Sigma^{-1} a_v'$  has the same value  $\tau'\tau$  for every  $v$ . Second, while Theorem 6.1 as stated includes all possible arrangements of  $h$  outliers as alternatives, the discussion applies if only an arbitrary subset of these arrangements are considered as alternatives. Modifying the interpretation of  $\Sigma_v$  to include only the specified permutations  $v$  is the only change needed for this result. Third, if  $a_v \Sigma^{-1} a_v'$  as well as  $a_v$  varies with  $v$ , but is known up to a constant, then the insertion of this known scalar term as a factor accompanying  $\tau'\tau$  leads without difficulty to a broader version of the theorem. In this situation, too, we can consider only some subset of  $h$ -outlier alternatives. These generalizations will not be stated as theorems, although they could be.

The simplest case of  $h$  outliers occurs in Theorem 6.1 when  $c_i = 1$  for  $i = 1, \dots, h$ , and  $\delta_v = 1$  for all permutations  $v$ . There are  $\binom{n}{h}$  distinct  $h$ -outlier matrices  $A$ , corresponding to all possible choices of the  $h$  identical non-zero rows of  $A$ . It simplifies the discussion if we now return to the notation  $H_{i_1, \dots, i_h}$  instead of  $H_v$ . If rows  $i_1, \dots, i_h$  of  $A$  are the non-zero rows, then the column vector  $I^v(c_1 \dots c_n)'$   $= [c_{v(1)} \dots c_{v(n)}]'$  has ones in those entries and zeroes elsewhere. Then by Lemma 6.2,  $s_v$  can be written as

$$(6.11) \quad (R_{i_1} + \dots + R_{i_h})(R'R)^{-1}(R_{i_1} + \dots + R_{i_h})'$$

Substitution of these expressions into Theorem 6.1 produces

Corollary 6.1: In the multivariate regression model with mean slippage, take  $h < n/2$ . Define alternative hypotheses  $H_0: A = 0$  and  $H_{i_1, \dots, i_h}$ :

rows  $i_1, \dots, i_h$  of  $A$  are all equal to  $a(1 \times p)$ , the remaining rows of  $A$  are zero. Let  $s_{i_1, \dots, i_h}$  denote the expression (6.11), and let  $\{f_{i_1, \dots, i_h}\}$  be an arbitrary set of  $\binom{n}{h}$  positive constants, and define

$$f_{\text{max}} = \max_{i_1, \dots, i_h} f_{i_1, \dots, i_h} s_{i_1, \dots, i_h}.$$

Consider the decision rule which states: take action  $D_0$  when  $f_{\text{max}} < K$ ; take action  $D_{i_1, \dots, i_h}$  when  $f_{i_1, \dots, i_h} s_{i_1, \dots, i_h} = f_{\text{max}} \geq K$ . The constant  $K$  is determined by the desired value of  $\Pr(D_0 | H_0)$ . If more than one collection  $i_1, \dots, i_h$  is associated with  $f_{\text{max}}$ , which happens with probability zero, the choice among the corresponding actions may be made in any pre-specified fashion, possibly at random. Any decision rule of this form is invariant admissible for this outlier problem. In fact, it is Bayes with respect to any prior distribution of the form

$$p_{i_1, \dots, i_h} = c_0 \alpha_{i_1, \dots, i_h}^{-1} \rho(s_{i_1, \dots, i_h}, f_{i_1, \dots, i_h}) / \rho(s_{i_1, \dots, i_h}, a \Sigma^{-1} a'),$$

$$p_0 = 1 - \sum p_{i_1, \dots, i_h},$$

where

$$\alpha_{i_1, \dots, i_h} = \exp \left[ -\frac{1}{2} a \Sigma^{-1} a' - \sum_{j, j' = i_1, \dots, i_h} m_{jj'} \right],$$

and  $c_0$  is chosen to make all of these positive.

The results of this section extend work of Murphy (1951) on univariate observations. His no-outlier null hypothesis is that the  $n$  observations are a random sample from a  $N(\mu, \sigma^2)$  population, and each  $h$ -outlier alternative is that exactly  $h$  specified observations come instead from a  $N(\mu + \lambda \sigma, \sigma^2)$  population, where  $\lambda > 0$ . Murphy's test rejects the null hypothesis for large values of the statistic

$$[x_{(n)} + x_{(n-1)} + \dots + x_{(n-h+1)} - h\bar{x}] / s,$$

which also selects the  $h$  outliers. The approach of this section is more general, treating multivariate observations, a regression model rather than a common mean model, and  $h$  outliers of differing magnitudes.

The invariance structure of this section is not the one Murphy adopted. Murphy, like Ferguson (1961), required his procedure to be invariant only under linear transformations of the data which do not reverse its direction:  $x_i \rightarrow ax_i + b$  with  $a > 0$ . In a multivariate context, it is more natural to require invariance under all non-singular linear transformations:  $x_i \rightarrow ax_i + b$  with  $a \neq 0$ . This corresponds to assuming  $\lambda \neq 0$  rather than  $\lambda > 0$  as above. The decision rule of Corollary 6.1 with all  $f$ 's = 1 is easily obtained for the univariate, common mean case of Murphy as: reject the null hypothesis for large values of

$$\max\{[x_{(n)} + \dots + x_{(n-h+1)} - h\bar{x}]/s, [h\bar{x} - x_{(1)} - \dots - x_{(h)}]/s\}.$$

This test is clearly preferable to Murphy's when the sign of  $\lambda$  is not known.

## Bibliography

- Anscombe, F.J. (1960). Rejection of outliers. Technometrics 2, 123-147.
- Anscombe, F.J. and Tukey, J.W. (1963). The examination and analysis of residuals. Technometrics 5, 141-160.
- Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. John Wiley and Sons, New York.
- Cook, R.D. (1977). Detection of influential observations in linear regression. Technometrics 19, 15-18.
- David, H.A. (1970). Order Statistics. John Wiley and Sons, New York.
- Dixon, W.J. (1950). Analysis of extreme values. Ann. Math. Statist. 21, 488-506.
- Doornbos, R. (1966). Slippage Tests. Mathematical Centre Tracts 15, Mathematisch Centrum, Amsterdam.
- Ferguson, T.S. (1961). On the rejection of outliers. Proc. Fourth Berkeley Symposium Math. Statist. Prob. I, 253-287.
- Ferguson, T.S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.
- Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. John Wiley and Sons, New York.
- Graybill, F.A. (1961). An Introduction to Linear Statistical Models, Volume I. McGraw-Hill Book Company, Inc., New York.
- Grubbs, F.E. (1950). Sample criteria for testing outlying observations. Ann. Math. Statist. 21, 27-58.
- Grubbs, F.E. (1969). Procedures for detecting outlying observations in samples. Technometrics 11, 1-21.
- Karlin, S. and Truax, D.R. (1960). Slippage problems. Ann. Math. Statist. 31, 296-324.
- Kruskal, W.H. (1960). Some remarks on wild observations. Technometrics 2, 1-3.
- Lehmann, E.L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, New York.
- Murphy, R.B. (1951). On tests for outlying observations. Ph.D. Thesis, Princeton University.
- Press, S.J. (1972). Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc., New York.